

---

# Anisotropic clustering analysis of the Quasar Sample from SDSS-IV extended Baryon Oscillation Spectroscopic Survey

Jiamin Hou

---



München 2020



---

# Anisotropic clustering analysis of the Quasar Sample from SDSS-IV extended Baryon Oscillation Spectroscopic Survey

Jiamin Hou

---

Dissertation  
an der Fakultät für Physik  
der Ludwig–Maximilians–Universität  
München

vorgelegt von  
Jiamin Hou  
aus Shanghai

München, den 4.5.2020

Erstgutachter: Prof. Dr. Ralf Bender

Zweitgutachter: Prof. Dr. Jochen Weller

Tag der mündlichen Prüfung: 25.6.2020



- *"flow of life emerges"*

- *"yes, I know", said Mian Octobers, "from collective consciousness".*



# Table of contents

<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The observable universe . . . . .	1
1.2 A review on the galaxy redshift surveys . . . . .	3
1.3 Overview of eBOSS survey . . . . .	6
1.4 Future surveys . . . . .	8
1.5 Outline of the thesis . . . . .	11
<b>2 Theory basics</b>	<b>13</b>
2.1 Einstein equations and general relativity . . . . .	13
2.1.1 Inhomogeneity in a smooth background . . . . .	15
2.2 Density perturbation in the early time – linear regime . . . . .	19
2.2.1 Equations of motion . . . . .	19
2.2.2 N-point statistics . . . . .	20
2.3 Modelling the observational LSS clustering . . . . .	21
2.3.1 BAO as a standard ruler . . . . .	21
2.3.2 RSD in linear regime . . . . .	23
<b>3 Modelling the 2-point correlation function</b>	<b>27</b>
3.1 Gravitational evolution in the nonlinear regime . . . . .	28
3.1.1 Standard perturbation theory . . . . .	28
3.1.2 Renormalized perturbation theory . . . . .	30
3.1.3 Galilean-Invariant RPT . . . . .	31
3.1.4 RESPRESSO . . . . .	31
3.1.5 Fitting function for the velocity . . . . .	33
3.2 Large scale structure bias . . . . .	35

3.2.1	Local and nonlocal bias . . . . .	36
3.2.2	Higher-derivative bias . . . . .	37
3.3	Nonlinear modelling of the RSD . . . . .	38
3.3.1	Nonlinear RSD on large scales . . . . .	38
3.3.2	Nonlinear RSD at small scale: Finger of God factor . . . . .	39
<b>4</b>	<b>Clustering analysis of the eBOSS QSO</b>	<b>43</b>
4.1	Data . . . . .	43
4.1.1	Overview of the eBOSS survey on QSOs . . . . .	43
4.1.2	Imaging phase . . . . .	44
4.1.3	Spectroscopic phase . . . . .	45
4.2	Two point correlation function . . . . .	50
4.3	Methodology . . . . .	53
4.3.1	Inference of cosmological parameters . . . . .	53
4.3.2	Covariance matrix estimation . . . . .	54
4.4	Assessing the systematic uncertainties . . . . .	58
4.4.1	Modelling systematics: non-blind mock challenge . . . . .	59
4.4.2	Fiducial cosmology systematics: blind mock challenge . . . . .	63
4.4.3	Observational systematics: EZMOCKS . . . . .	66
4.5	BAO-only analysis . . . . .	72
<b>5</b>	<b>Cosmological implication from eBOSS</b>	<b>75</b>
5.1	Constraints from the eBOSS QSOs DR16 QSO analysis . . . . .	75
5.1.1	Results in the configuration space: full-shape analysis . . . . .	75
5.1.2	Results in the configuration space: BAO-only analysis . . . . .	79
5.1.3	Combination of the configuration space and Fourier space results . . . . .	82
5.2	Robustness tests on the data analysis . . . . .	85
5.2.1	List of tests performed on the data . . . . .	86
5.2.2	Summary of the robustness test . . . . .	89
<b>6</b>	<b>RSD and the modified gravity models</b>	<b>91</b>
6.1	Introduction to modified gravity models . . . . .	91
6.2	Structure formation in modified gravity theory . . . . .	93
6.2.1	$f(R)$ gravity . . . . .	93
6.2.2	Dvali–Gabadadze–Porrati model . . . . .	94
6.2.3	Screening mechanism . . . . .	95
6.3	Construction of the mock catalogues . . . . .	96

---

6.4	Fitting template for two point correlation . . . . .	97
6.5	Results and discussion . . . . .	98
<b>7</b>	<b>Summary and outlook</b>	<b>103</b>
7.1	Clustering analysis of the eBOSS QSO sample . . . . .	103
7.2	RSD in modified gravity theories . . . . .	105
7.3	Future perspective . . . . .	106
	<b>References</b>	<b>109</b>
	<b>Appendix A BAO reconstruction of the QSO catalog</b>	<b>119</b>
	<b>Appendix B Supplementary to the eBOSS clustering analysis</b>	<b>123</b>
B.1	Results expressed in alternative basis . . . . .	123
B.1.1	$D_M/r_{\text{drag}}-Hr_{\text{drag}}-f\sigma_8$ basis . . . . .	123
B.2	Parameter settings for the OuterRim mock challenge . . . . .	124



# List of figures

1.1	Map of galaxy counts in the southern galactic hemisphere from Lick galaxy catalog. . . . .	3
1.2	History of the BAO distance measurement. . . . .	5
1.3	SDSS plate and a 3D representation of the distant objects . . . . .	7
1.4	Constraints on $\gamma$ and $w$ from Euclid survey. . . . .	9
1.5	Progress in the galaxy redshift survey over the decades. . . . .	10
2.1	Schematic drawing of AP test. . . . .	23
3.1	Comparison of RESPRESSO matter power spectrum. . . . .	33
3.2	Comparison of cross power spectrum for the matter-velocity divergence $P_{\delta\theta}$ and the auto power spectrum for the velocity divergence $P_{\theta\theta}$ . . . . .	35
3.3	PDF for the pairwise velocity along the line of sight direction in OUTERRIM mocks. . . . .	41
4.1	Imaging systematics for DR16 QSO sample. . . . .	46
4.2	DR16 QSO footprints. . . . .	47
4.3	Comparison for DR14 and DR16 QSO sample. . . . .	47
4.4	Number density for DR16 QSO sample. . . . .	48
4.5	Fiber efficiency as a function of fiber ID. . . . .	48
4.6	2-dimensional correlation function $\xi(s_{\perp}, s_{\parallel})$ for DR16 QSO sample. . . . .	51
4.7	Two-point correlation function for DR14 analysis . . . . .	52
4.8	Two-point correlation function for DR16 analysis . . . . .	53
4.9	Error on the inferred parameters as a function of number of mock simulations	56
4.10	Contours of the two-dimensional correlation function from OUTERRIM mocks.	61
4.11	Constraints on AP parameters and growth rate parameter inferred from four redshift distribution schemes for OUTERRIM HODs mocks. . . . .	62
4.12	Redshift uncertainty parameter as a function of the satellite fraction. . . . .	64

4.13	Multipoles measured from EZMCKS with different combination of systematics. . . . .	69
4.14	Distribution of the radial number density in EZMOCKS. . . . .	70
4.15	Projected correlation function in EZMOCKS. . . . .	71
5.1	Posterior distribution of the AP parameters and the structure growth parameter inferred from the DR16 QSO sample. . . . .	76
5.2	Distribution of statistical error for the mocks and data. . . . .	77
5.3	Redshift evolution of the cosmological distance measurements and growth rate compared to $\Lambda$ CDM + Planck. . . . .	78
5.4	Likelihood of the BAO parameters. . . . .	80
5.5	Comparison of full shape (red contour) and BAO only (blue contour) fits in configuration space. . . . .	82
5.6	Correlation coefficients for the configuration multipoles and the power spectrum. . . . .	83
5.7	Posterior distribution for $\alpha_{\perp}$ , $\alpha_{\parallel}$ and $f\sigma_{12}$ for the combined configuration space and Fourier space analysis. . . . .	84
5.8	Projected correlation function of DR16 QSO sample. . . . .	88
6.1	Clustering multipoles and wedges for different modified gravity models. . .	99
6.2	RSD distortion parameter $\beta(z)$ for multipoles and wedges with $s_{\min} = 20, 40 h^{-1}\text{Mpc}$ . . . . .	100
A.1	A slice of the BAO reconstructed density field projected along x-axis of the simulation box. . . . .	120
A.2	Two point correlation from reconstructed OUTERRIM mocks . . . . .	120
A.3	Two point correlation from the reconstructed DR16 QSO catalog . . . . .	121



# List of tables

4.1	Summary of properties of the eBOSS DR16 QSOs clustering catalogue. . .	44
4.2	Statistics obtained from OUTERRIM nonblind mock challenge. . . . .	64
4.3	Statistics obtained the OUTERRIM blind mock challenge with fixed HODs. . . . .	67
4.4	Summary of systematic error for the OUTERRIM mock challenge with different HODs. . . . .	67
4.5	Summary of observational systematics measured from the EZMOCKS. . . . .	73
5.1	Summary of the parameter space $\lambda$ . . . . .	79
5.2	Final systematics for full-shape and BAO-only fit to the mocks. . . . .	81
5.3	Constraints on the DR16 data in configuration space, Fourier space and combined results. . . . .	84
5.4	Systematics from the OUTERRIM mock challenge of the combined statistics in configuration and Fourier space. . . . .	86
5.5	Effective redshift with different definitions. . . . .	89
5.6	Summary of the robustness test on the DR16 QSO sample. . . . .	90
B.1	HOD parameters for OUTERRIM nonblind mocks. . . . .	125
B.2	Parameter settings for OUTERRIM blind mock challenge. . . . .	126

# Zusammenfassung

Die moderne Kosmologie bietet eine Beschreibung der thermischen Geschichte des Universums sowie seiner Zusammensetzung. Der dominierende Energiegehalt des Universums ist eine Komponente namens *Dunkler Energie*, die die Beschleunigung der kosmischen Expansion zu später Zeit bewirkt. Mehrere Datensätze, wie die der kosmischen Mikrowellen-Hintergrundexperimente (MWH) und Typ Ia Supernovae, halfen uns, ein Standardparadigma für das  $\Lambda$ CDM-Modell aufzustellen. Die Natur der dunklen Energie ist dadurch jedoch noch nicht festgelegt. Das Verständnis der Materieverteilung zu einem späten kosmischen Zeitpunkt ist ein mächtiges Werkzeug, das uns strenge Einschränkungen für diese mysteriöse Komponente bietet.

Diese Arbeit präsentiert die Messungen der besonders ehrgeizigen Himmelsbeobachtung Sloan Digital Sky Survey (SDSS), der die ausführlichste dreidimensionale Kartierung des Universums erstellt hat. Der extended Baryon Oscillation Spectroscopic Survey (eBOSS), als Unterprogramm des SDSS IV-Projekts zielte darauf ab, die Expansionsgeschichte des Universums bis zu einem Zeitpunkt abzubilden, an dem es weniger als drei Milliarden Jahre alt war. Der vollständige eBOSS-Datensatz deckt über achtzig Prozent der kosmischen Geschichte ab.

Bisher wurden Galaxien als Tracer für das zugrunde liegende Materiefeld verwendet. Man misst die Galaxienspektren, die durch die kosmische Expansion "gedehnt und rotverschoben" werden. Die Rotverschiebung steht in direktem Zusammenhang mit ihrer Entfernung zu den Beobachtern. In dieser Arbeit konzentrieren wir uns auf die quasi-stellaren Objekte (QSOs). Es wird angenommen, dass QSOs von supermassiven Schwarzen Löchern angetrieben werden. Daher sind sie an sich leuchtender als die Galaxien und ermöglichen es uns, den messbaren Rotverschiebungsbereich zu erweitern. Wir untersuchen die Clustering-Eigenschaften der QSOs und verwenden *Baryonische akustische Oszillationen* (BAO) als Distanzlineal, um unser Wissen über das  $\Lambda$ CDM-Modell zu überprüfen. Darüber hinaus haben beobachtete Rotverschiebungen eine zusätzliche Komponente, einen kleinen Beitrag aufgrund der Eigengeschwindigkeit der Galaxie als Reaktion auf die gravitative Anziehung der umgebenden Materie. Man kann diese zusätzliche Komponente, bekannt als *redshift space distortion* (RSD), verwenden, um das Wachstum der kosmischen Struktur zu untersuchen.

Die eBOSS DR16 QSO-Probe enthält 343,708 spektroskopisch bestätigte Quasare mit einer Rotverschiebung  $0.8 < z < 2.2$ . Bei einer effektiven Rotverschiebung von  $z_{\text{eff}} = 1.48$  messen wir die Winkelentfernungsfunktion  $D_M(z_{\text{eff}})/r_{\text{drag}} = 30.66 \pm 0.88$ , die Hubble-Distanz  $D_H(z_{\text{eff}})/r_{\text{drag}} = 13.11 \pm 0.52$ , und die Wachstumsrate  $f\sigma_{12}(z_{\text{eff}}) = 0.435 \pm 0.048$ . Die Genauigkeit dieser Messungen wird durch umfangreiche Mehrkörpersimulationen bestätigt, die für die Quasarprobe entwickelt wurden, sowie durch die Untersuchung der realistischen systematischen Beobachtungsfehler. Wir führen eine reine BAO-Analyse durch, um die Robustheit der Methodik der Vollformanalyse zu überprüfen. Wenn wir unsere Analyse mit der Fourier-Raum-Analyse kombinieren, erhalten wir  $D_M^c(z_{\text{eff}})/r_{\text{drag}} = 30.21 \pm 0.79$ ,  $D_H^c(z_{\text{eff}})/r_{\text{drag}} = 13.23 \pm 0.47$ , und  $f\sigma_{12}^c(z_{\text{eff}}) = 0.458 \pm 0.045$ . Die kombinierte Abstandsmessung stimmt hervorragend mit dem  $\Lambda$ CDM-Modell mit den bestpassenden kosmologischen Parameterwerten aus CMB-Planck-Daten überein. Der Wachstumsratenparameter zeigt eine Abweichung von  $\sim 2\sigma$  im Vergleich zu den Vorhersagen des  $\Lambda$ CDM-Modells.

In dieser Arbeit zeigen wir auch, wie die RSD-Messungen verwendet werden können, um mögliche Abweichungen von der Vorhersage der allgemeinen Relativitätstheorie zu testen und eine mögliche alternative Lösung für das  $\Lambda$ CDM-Modell bereitzustellen. Wir verwenden synthetische Galaxienkataloge, die aus großen N-Körpersimulationen von Standardmodellen und modifizierten Gravitationsmodellen erstellt wurden, um das Galaxien-Clustering im Rotverschiebungsraum zu messen. Wir überprüfen zwei repräsentative Familien modifizierter Gravitationsmodellen: das Hu & Sawicki  $f(R)$ -Modell und der normale Zweig des Dvali-Gabadadze-Porrati (DGP)-Modells. Wir haben bestätigt, dass die Standardanalyse-Pipeline auf die modifizierten Schwerkraftmodelle anwendbar ist. Darüber hinaus zeigt der RSD-Verzerrungsparameter Abweichungen von der Vorhersage der Standardgravitation in den DGP-Modellen, die für zukünftigen Surveys interessant sein werden.

# Abstract

Modern cosmology offers a description of the thermal history of the Universe as well as its composition. The dominant energy content of the Universe appears to be a component called *dark energy*, which sources the acceleration of the cosmic expansion at late time. Multiple data sets, such as the cosmic microwave background experiments (CMB), type Ia supernovae, helped us to set up a standard paradigm of the  $\Lambda$ CDM model. However, it does not yet specify the nature of the dark energy. The study of the matter distribution at late cosmic time is a powerful tool that provides us with powerful constraints on this mysterious component.

This thesis presents the measurements from one of the most ambitious sky observations from the Sloan Digital Sky Survey (SDSS), which has created the most detailed three-dimensional maps of the Universe. The extended Baryon Oscillation Spectroscopic Survey (eBOSS), as a subprogram from the SDSS IV project, aimed at mapping out the expansion history of the Universe back to the point when it was less than three billion years old. The complete eBOSS data set covers over eighty percent of cosmic history.

Previously, galaxies have been used as tracers of the underlying matter field. One measures the galaxies spectra, which are “stretched and reddened” by the cosmic expansion. The redshift is directly related to its distance to the observers. In this thesis we focus on the quasi-stellar objects (QSOs). QSOs are believed to be powered by supermassive black holes, therefore, they are intrinsically more luminous than the galaxies and allow us to extend the redshift range. We study the clustering properties of the QSOs and use the *baryon acoustic oscillation* (BAO) as a distance ruler to verify our knowledge of the  $\Lambda$ CDM model. Furthermore, observed redshifts have an additional component, a small contribution due to the galaxy’s own velocity in response to the gravitational attraction of the surrounding matter. One can make use of this additional component, known as *redshift space distortion* (RSD), to study the growth of the cosmic structure.

The eBOSS DR16 QSO sample contains 343,708 spectroscopically confirmed quasars between redshift  $0.8 < z < 2.2$ . At an effective redshift of  $z_{\text{eff}} = 1.48$ , we measure the comoving angular diameter distance  $D_M(z_{\text{eff}})/r_{\text{drag}} = 30.66 \pm 0.88$ , the Hubble distance  $D_H(z_{\text{eff}})/r_{\text{drag}} = 13.11 \pm 0.52$ , and the growth rate  $f\sigma_{12}(z_{\text{eff}}) = 0.435 \pm 0.048$ . The accuracy of these measurements is confirmed using an extensive set of mock simulations developed

for the quasar sample, we also examined various realistic observational systematics. We perform a BAO-only analysis to cross check the robustness of the methodology of the full-shape analysis. Combining our analysis with the Fourier space analysis, we arrive at  $D_M^c(z_{\text{eff}})/r_{\text{drag}} = 30.21 \pm 0.79$ ,  $D_H^c(z_{\text{eff}})/r_{\text{drag}} = 13.23 \pm 0.47$ , and  $f\sigma_{12}^c(z_{\text{eff}}) = 0.458 \pm 0.045$ . The combined distance measurement is in excellent agreement with the  $\Lambda$ CDM model with best fitting values from CMB Planck data. The growth rate parameter shows a  $\sim 2\sigma$  deviation compared to the  $\Lambda$ CDM prediction.

In this thesis we also demonstrate how RSD measurements can be used to test for potential deviations from the prediction of general relativity and provide potential alternative solution to the  $\Lambda$ CDM model. We use mock galaxy catalogues constructed from large N-body simulations of standard and modified gravity models to measure galaxy clustering in redshift space. We check two representative families of modified gravity models: Hu & Sawicki  $f(R)$  model and the normal branch of the Davli-Gabadadze-Porrati (DGP) model. We confirmed that the standard analysis pipeline is applicable to the modified gravity models. Furthermore, RSD distortion parameter shows deviations from the prediction of standard gravity in the DGP models, which will be interesting to explore in the future survey.



# Chapter 1

## Introduction

### 1.1 The observable universe

The modern standard cosmological model provides a thermal history as well as the energy content of the Universe.

In the standard cosmological picture, the Universe began with a hot Big Bang. Very shortly after the Big Bang, a rapid inflationary phase near the energy scale ( $\gtrsim 10^{16}$  GeV) took place and the Universe expanded by a factor of  $\sim 10^{26}$  during this period. A *Grand Unification* of the electroweak and strong interaction were expected at such high energy scale. The Universe cooled down while expanding, when the temperature dropped down to  $\sim 100$ - $10^3$  GeV, the electroweak and strong interaction in the standard particle model start to be applicable. As the temperature reached  $\sim 100$  MeV, quarks and gluons first became confined within neutrons and protons. The primordial neutrinos decoupled from other particles when the temperature dropped down to  $\sim 1$  MeV. A direct consequence of the neutrino decoupling is that the chemical equilibrium between neutrons and protons via the weak interaction stopped and their ratio remain frozen. Further, at energy scale of  $\sim 0.05$  MeV, the primordial nucleosynthesis took place. Neutrons and protons combined to form helium and other light elements. All these processes happened only within around 3 minutes since the begin of the Universe. After the temperature arrived at  $\sim 1$  eV, the Universe was dominated by the radiation pressure, the photons and relativistic electrons interacted via Compton scattering and re-distribute energy and momentum<sup>1</sup>, while electrons and protons were coupled via Coulomb force. They formed the *baryon-photon plasma*. As the Universe further cooled down, the photon finally decoupled from this plasma and form the *cosmic microwave background* (CMB).

---

<sup>1</sup>Compton scattering reduces to Thomson scattering at low energy.

More than one hundred years ago, it was already realised that the Universe has been expanding (via observing the shift of spectral lines of spiral galaxies, 1912, by Vesto Slipher ([Slipher, 1914](#))). Later on, the study was followed by Edwin Hubble who proposed the well known Hubble's law  $v = H_0 D$ , stating that the recession velocity  $v$  of the objects is proportional to the Hubble parameter and the distance of the objects to the observer ([Hubble, 1929](#)). What's more, in the late 90's, the study on Type Ia supernovae ([Perlmutter et al., 1999](#); [Riess et al., 1998](#)) implied that the Universe is currently experiencing an accelerated expansion. An unknown component *dark energy* was believed to be the source of the accelerated expansion.

The Universe is full of surprises. Back in 1930's, the existence of *dark matter* was posited by Jan Oort and Fritz Zwicky as they studied the orbital/rotational speeds of stars/galaxies ([Oort, 1932](#); [Zwicky, 1933](#)). Later on, multiple observational evidences supported the hypothesis of the dark matter, including the gravitational lensing ([Taylor et al., 1998](#)), galaxy clusters ([Clowe et al., 2004](#); [Markevitch et al., 2004](#)), and most remarkably, the tiny temperature fluctuation in the CMB experiments. In the baryon-photon plasma mentioned above, the density perturbations experienced two counteracting forces: photon pressure and the gravitational interaction. These two effects give rise to acoustic waves that formed the *baryon acoustic oscillation* (BAO). In order to describe the temperature power spectrum with a series of acoustic peaks measured with high precision (see figure 1 in [Collaboration et al. \(2016\)](#)), one must introduce the cold dark matter (CDM), where "cold" implies the particles being massive and non-relativistic. The power spectrum with rich feature can be well fit by the  $\Lambda$ CDM model and offer tight constraints on the energy budget of the dark energy, dark matter, the baryonic matter (such as, stars, gases), and radiation (photons, relativistic neutrinos).

The modern cosmological model provides us a beautiful picture of how the Universe evolved and what is it composed of. However, it does not describe the nature of the components that dominate the energy content of the Universe. To further explore these questions, it is also very useful to look at the matter distribution at late time. The matter distribution traced by luminous galaxies can be used to predict the expansion history of the Universe and the growth of cosmic structure  $f$  (see Chapter 2). It can offer a clue of whether the dark energy is a cosmological constant with a static equation of state parameter  $w^2$ , or a scalar field (quintessence), or other possibilities. The growth rate of the cosmic structure can be used to test the possible deviation from the general relativity (GR), and potentially open another angle of viewing the origin of the cosmic acceleration.

---

<sup>2</sup> $w = p/\rho$ , with  $p$  being pressure and  $\rho$  being energy density.



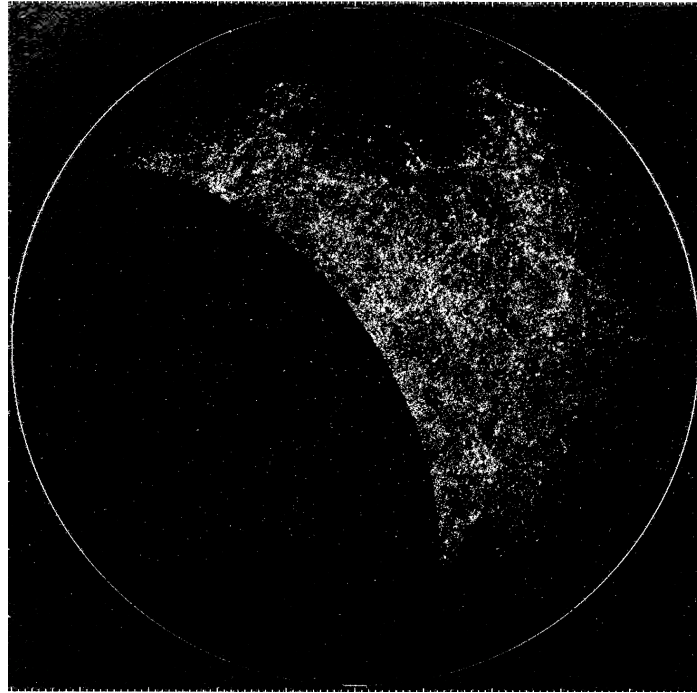


Fig. 1.1 Map of galaxy counts in the southern galactic hemisphere from Lick galaxy catalog. Galactic latitude is a linear function of radius and increases in the counter-clockwise direction from  $0^\circ$  at the bottom of the image. Image credit: [Seldner et al. \(1977\)](#)

This thesis is therefore dedicated to the clustering analysis of the large-scale structure (LSS) of the Universe.

## 1.2 A review on the galaxy redshift surveys

The measurement of the distribution of the LSS can be dated back to the 1960s'. The Lick galaxy catalog produced by [Shane & Wirtanen \(1967\)](#) recorded roughly a million galaxies using the refracting telescope at Lick Observatory using photographic plates. Fig. 1.1 shows maps of the corrected galaxy counts in the southern galactic hemispheres. In this catalogue, one can see the filaments of the cosmic web and it was the first time the structures of a few tens of Mpc are recorded. These catalogues allow us to study the angular projected-distribution of the galaxies.

The redshift surveys provide a 3D mapping of the objects in a field of the sky with the radial distribution of the galactic objects. The pioneering work done by the Center for Astrophysics (CfA) Redshift Survey was the first attempt to map the large-scale structure of the universe ([Huchra et al., 1983](#)). The original CfA survey contained redshifts for 2,400 galaxies, where one can identify the cosmic structures such as voids and galaxy clusters. It

has been known for long time that the peculiar velocity of the galaxies can cause an angular dependent clustering signal, known as the redshift space distortion (Jackson, 1972; Kaiser, 1987). Observationally, it was first demonstrated using the PSCz catalogue (Saunders et al., 2000) that RSD has an impact on the galaxy density field in the redshift space. Tadros et al. (1999) successfully determined the distortion parameter that quantifies the degree of the anisotropy in the galaxy clustering.

The spectroscopic surveys grew rapidly. The Two-degree-Field Galaxy Redshift Survey (2dF-GRS; Colless et al., 2001) covered  $2000 \text{ deg}^2$  and contained redshifts for  $\sim 250,000$  galaxies. The survey operated from 1997-2003 allowed accurate measurements of the power spectra. Fits to the power spectra data provided a measurement of the matter content of the Universe (Cole et al., 2005). The joint likelihood analysis of the power spectra with a compilation the cosmic microwave background (CMB) observation provided further constraints on the cosmological model as wells as the cosmological parameters (Efstathiou et al., 2002; Sánchez et al., 2006). 2dF-GRS was also the first survey measured the galaxy bias (Lahav et al., 2002) and performed the detailed analysis of the distortion of the clustering pattern in redshift space (Peacock et al., 2001).

Later on the Sloan Digital Sky Survey (SDSS) (York et al., 2000) overtook the place of 2dF-GRS and became the largest galaxy redshift survey. During its first two phases SDSS-I/SDSS-II (2000–2008), the survey already imaged over  $8,000 \text{ deg}^2$  of the sky in five optical bandpasses. It imaged more than one million galaxies and 100,000 quasars, among which, 260,490 were spectroscopically confirmed. In 2005, the first detection of BAO in the configuration space was published by Eisenstein et al. (2005) using a sample of 46,748 luminous red galaxies (LRG) optimized for LSS clustering study. This finding was complemented by the power spectrum analysis of the 2dF-GRS (Cole et al., 2005). In the SDSS-II final data release DR7, the volume-averaged cosmological distance  $D_V^3$  can be constrained to a few percent level (Reid et al., 2010). Ever since BAO became a standard probe of the cosmic distance measurements. The redshift galaxy surveys confirmed the discovery by the supernovae distance measurement (Riess et al., 1998) that the universe is undergoing an accelerated expansion, where a mysterious component known as dark energy needs to be introduced. These observations established the current standard model of cosmology, the  $\Lambda$ CDM model, where the equation of state parameter  $w = p/\rho$  for the dark energy is a constant over the cosmic time, with  $p$  being pressure and  $\rho$  being energy density.

During the third phase, SDSS started a 5-year program, the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al., 2013). BOSS performed spectroscopic observations

---

<sup>3</sup>Volume-averaged cosmological distance  $D_V(z) = \left( D_M^2 \frac{cz}{H(z)} \right)^{\frac{1}{3}}$ , as a combination of the comoving angular diameter distance  $D_M(z)$  and the Hubble parameter  $H(z)$

on more than 1.5 million galaxies and 150,000 quasars. Over 5-years the program covered  $10,000 \text{ deg}^2$  and measured the cosmic distance using BAO for redshift  $z < 0.6$  and  $z = 2.5$ . During this phase, more complicated models were available in describing both the BAO signal and the RSD pattern. Fig. 1.2 shows the redshift evolution of the volume-averaged distance  $D_V(z)$ . The solid curve is the ratio of the volume-averaged distance predicted using  $\Lambda\text{CDM}$  model with input best-fit cosmological parameters from Planck-2018 (Collaboration et al., 2018) to the fiducial cosmology adopted by the BOSS program. The constraints obtained from BOSS are the three grey points denoted by “BOSS-LRG” (Alam et al., 2017) that used luminous red galaxies as tracers and the cyan points denoted by “BOSS- $\text{Ly}\alpha$ ” (Bautista et al., 2017) that used  $\text{Ly}\alpha$  forests as tracers. The other points were from the pre-BOSS measurement. The BOSS program provided much tighter constraints on the distance measurement at low redshift range compared to the previous surveys. At the same time, it also provided a tracer at high redshift extend beyond  $z > 2$ .

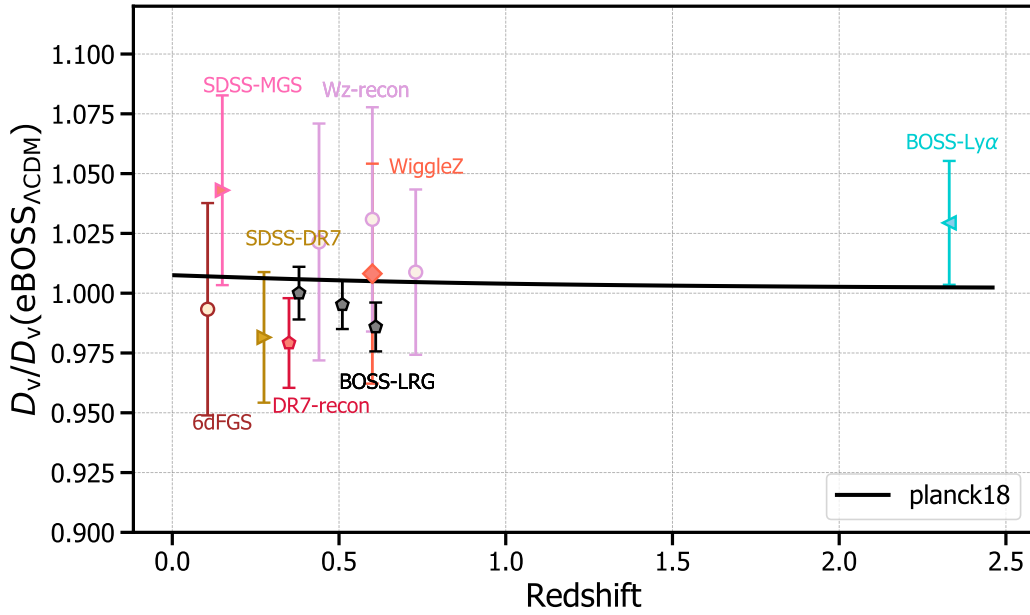


Fig. 1.2 Volume-averaged  $D_V$  distance as a function of redshift. The solid curve is the ratio of  $D_V$  predicted using  $\Lambda\text{CDM}$  model with input best-fit cosmological parameters from Planck-2018 (Collaboration et al., 2018) to the fiducial cosmology adopted by the BOSS program. The three grey points denoted by “BOSS-LRG” (Alam et al., 2017) use luminous red galaxies as tracers, and the cyan points denoted by “BOSS- $\text{Ly}\alpha$ ” (Bautista et al., 2017) that use  $\text{Ly}\alpha$  forests as tracers. The other points are obtained from the pre-BOSS measurements.

SDSS-IV is the current phase of the survey (2014-2020) (Blanton et al., 2017) with the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) being the cosmological program. The past surveys successfully established/confirmed the standard cosmological

model and confined constraints on the spatial curvature of the Universe as well as the cosmic densities for the baryonic matter, cold dark matter. However, the  $\Lambda$ CDM model with a constant equation of state parameter  $w = -1$  still lacks a physical motivation, and the nature of the dark energy is still unknown. There exist other possibilities such as modified model of gravity, dynamic fields, new fluids, etc. that offer alternative explanations to the standard picture. In order to verify those models, future measurements at different redshift bins that map the expansion history of the Universe to a higher precision are very useful.

### 1.3 Overview of eBOSS survey

The extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al., 2016) program aims at a precise measurement of the expansion history of the Universe to a critical early cosmic phase and improve constraints on the nature of dark energy. The eBOSS program was commenced in July 2014 and is performed during the fourth phase SDSS-IV. The eBOSS program used the 2.5-meter SDSS telescope (located at APO in New Mexico) (Gunn et al., 2006) but inherited the double-armed BOSS spectrographs (Smee et al., 2013). These spectrographs were built with smaller fibers, improved detectors and wider wavelength range 361 – 1014nm compared to the SDSS spectrograph. The spectrographs are fed by 1000 optical fibers, split into 500 for each, where each of the fiber subtends  $2''$  diameter in the sky. Fig. 1.3a shows a modern SDSS aluminium plate, drilled with tiny holes of  $9\mu\text{m}$  precision. Each hole is manually plugged with an optical fiber and each of them corresponds to an object in the sky. During each exposure optical fibers carry light from holes to a spectrograph. Fig 1.3b is a 3D representation of the distances to these objects with green being galaxies and purple being quasars.

There are four main tracers in the eBOSS program: luminous red galaxies (LRGs,  $0.6 < z < 0.8$ ), emission line galaxies (ELGs,  $0.7 < z < 1.1$ ), quasars for studying clustering (denoted as LSS quasar,  $0.8 < z < 2.2$ ), and another quasar sample for studies of the Ly $\alpha$  forest ( $z > 2.1$ ). They together cover a wide redshift range. The data collection of the eBOSS program, which ended on 1st March 2019, has undergone the data release 14 (DR14) in 2017, and is scheduled for data release 16 (DR16). The LSS quasar sample is the focus of this thesis, and we will present the data analysis of the final sample DR16. This quasar sample bridges the gap (as seen in Fig. 1.2) between lower redshift SDSS galaxy measurements (Alam et al., 2017; Beutler et al., 2011; Kazin et al., 2014) and those from the Ly $\alpha$ -forest (Bautista et al., 2017; du Mas des Bourboux et al., 2017).

The results presented in this thesis have been mainly published in Hou et al. (2018) and Hou et al. (2020). In the meantime, our analyses have significant contributions to many other

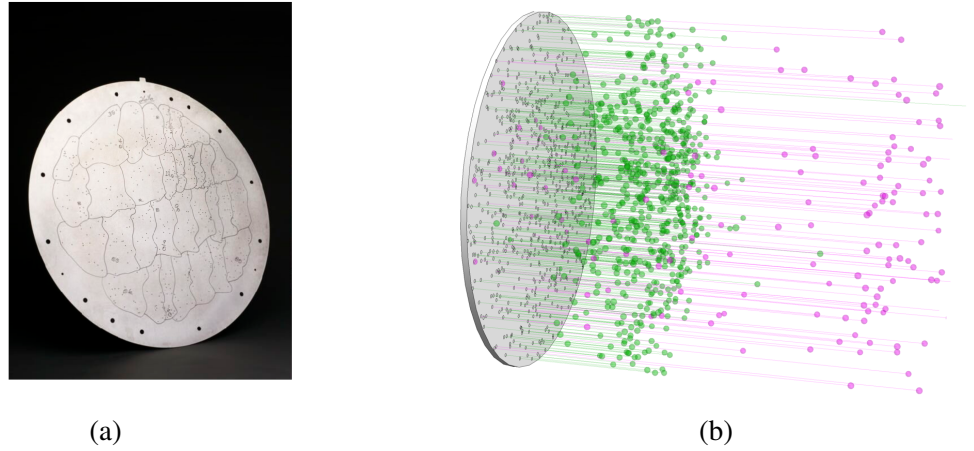


Fig. 1.3 The left plot shows the SDSS aluminum plate with tiny holes drilled (image credit: Astrophysical Research Consortium). The right plot shows a 3D representation of the distance to the objects with galaxies in green and quasars in purple (image credit: <https://blog.sdss.org> by David Kirkby).

eBOSS papers (Bautista et al., 2020; Gil-Marín et al., 2018, 2020; Lyke et al., 2020; Müller et al., 2020; Neveux et al., 2020; Ross et al., 2020; Smith et al., 2020; Tamone et al., 2020; Zarrouk et al., 2018; Zhao et al., 2019, 2020; de Mattia et al., 2020; du Mas des Bourboux et al., 2020).

In the data release DR14, Ata et al. (2017) first demonstrated the first detection of BAO using quasars as tracers. We published the results of the anisotropic clustering of the eBOSS QSO sample based on a smaller data sample using  $\sim 147,000$  quasars for redshift range at  $0.8 < z < 2.2$  (Hou et al., 2018). We decomposed the clustering signal into clustering wedges and the Legendre multipoles and used a state-of-the-art model for describing the clustering signal, the galileon invariant renormalized perturbation theory (gRPT; Crocce et al., prep) and modified the model to include realistic redshift distribution effect. As a joint effort, Zarrouk et al. (2018) presented analyses using a different model, Convolution Lagrangian Perturbation Theory (CLPT, Carlson et al., 2013) and the Gaussian streaming RSD model (Reid & White, 2011), in configuration space. Gil-Marín et al. (2018) performed analyses in Fourier space and used resummed perturbation theory calculation up to 2-loop order with the RSD model by Taruya et al. (2010). Ruggeri et al. (2019) optimised the measurements by applying a redshift-dependent weighting scheme using three Legendre multipoles in Fourier space and Zhao et al. (2019) applied RSD redshift weighting in Fourier space using two Legendre multipoles, with a power spectrum template calculated based on regularised perturbation theory up to second order (Taruya et al., 2012).

In the final data release DR16, we published the results based on a larger sample 343,708 objects (Hou et al., 2020). Compared to DR14, we applied new recipe in modelling the clustering signal and paid extensive effort in understanding various modelling systematics as well as the observational systematics. Afterwards, the model is modified to incorporate these new effects that were not studied deep enough in DR14. Ross et al. (2020) presented the clustering catalogue used for this analysis and the the complete DR16 quasar catalogue was described in Lyke et al. (2020). The quasar mock challenge used to assess modelling systematics was described in Smith et al. (2020). The approximate mocks used to estimate the covariance matrix and assess the observational systematics were presented in Zhao et al. (2020). A complementary quasar clustering analysis in Fourier space was performed in Neveux et al. (2020).

The framework of the DR16 analyses is structured similarly for LRG and ELG samples. For LRG sample, the catalogue was summarized in Ross et al. (2020). Gil-Marín et al. (2020) and Bautista et al. (2020) performed analysis in Fourier space and configuration space, respectively. Rossi et al. (2020) described the LRG mock challenge. For ELG sample, the catalogue was described in Raichoor et al. (2020), while Tamone et al. (2020) and de Mattia & Ruhlmann-Kleider (2019) performed analysis in configuration space and Fourier space, respectively. Alam et al. (2020) described the ELG mock challenge.

The BAO and RSD analyses of the QSO sample (Hou et al., 2020; Neveux et al., 2020), LRG sample (Bautista et al., 2020; Gil-Marín et al., 2020), ELG sample (Tamone et al., 2020; de Mattia et al., 2020) together with the BAO analyses of  $\text{Ly}\alpha$  (du Mas des Bourboux et al., 2020) entered Müller et al. (2020) for the cosmological implications from eBOSS.

## 1.4 Future surveys

Understanding the nature of dark energy is one of the main goals of the upcoming stage-IV experiments. We also face the puzzle in the Hubble tension between the experiments using the early universe information and the experiments measured in the local universe. The tension at order of  $\sim 4.4\sigma$  (Riess et al., 2019) implies room for alternative theories such as early dark energy components, screened fifth force (Desmond et al., 2018) and so on. The stage-IV experiments will map out larger volumes with increased statistical power, therefore we will have access to data with higher statistical precision.

The space-based Euclid mission (Amendola et al., 2018; Laureijs et al., 2011) is one of the leading stage-IV experiments. The most direct way to search for a deviation from  $\Lambda$ CDM model is to look for an equation of state parameter that is different from  $w \neq -1$ . Any deviation from  $w = -1$  would imply a non cosmological constant scenario. The Euclid



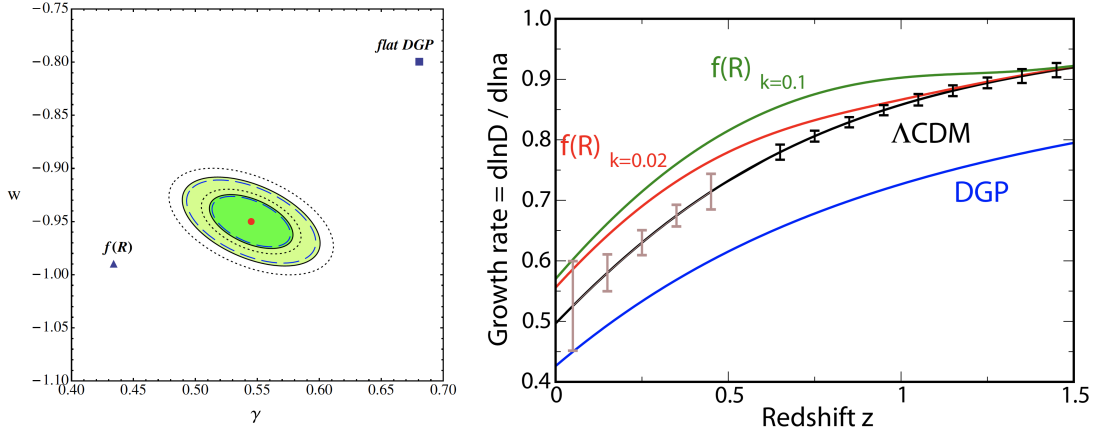


Fig. 1.4 Left panel: Marginalized probability of  $1\sigma$  and  $2\sigma$  regions for constant  $\gamma$  and  $w$ . The red dot marks the center for fiducial  $\Lambda$ CDM cosmology. The contours denote different error estimation on the bias and growth rate. The two alternative models  $f(R)$  (triangle) and flat DGP (square) can be clearly distinguished (Image credit: [Amendola et al., 2018](#)). Right panel: DESI forecast on the linear growth rate as a function of redshift. Two modified models are considered here: DGP model (blue) and  $f(R)$  model (whose scale- dependent growth we show evaluated at two different scales). Error bars are inferred from survey with DESI specifications with the brown (light) error bars at  $z < 0.5$  correspond to DESI bright galaxy survey (Image credit: [DESI Collaboration et al., 2016a](#))

mission will observe  $15,000 \text{ deg}^2$  that cover 10 billion sources. Among them, about 30 million sources will be used for galaxy clustering study. Fig. 1.4 shows the  $1\sigma$  and  $2\sigma$  for marginalized probability regions for constant  $\gamma$  and  $w$ , where  $\gamma(z)$  parametrizes the linear growth rate ([Linder & Cahn, 2007](#); [Wu et al., 2009](#)) (which will be further discussed in Section 2.3.2). The red dot marks the center of the fiducial  $\Lambda$ CDM model. Two modified gravity models,  $f(R)$  (add a function of Ricci scalar to the Einstein-Hilbert action) and Dvali-Gabadadze-Porrati (DGP; [Dvali et al., 2000](#)) in flat spacetime ([Maartens & Majerotto, 2006](#)), are plotted in triangle and square marker. In this plot, even with a pessimistic way of estimating the errors for the bias and growth rate (black dotted contour), one can distinguish these alternative models in a Euclid-like survey.

Another leading stage-IV spectroscopic experiment is the ground-based Dark Energy Spectroscopic Instrument (DESI; [DESI Collaboration et al., 2016a,b](#)). DESI will study the BAO and RSD with a wide area galaxy and quasar survey. The survey will extend the area to  $14,000 \text{ deg}^2$  and obtain 30 million galaxy and quasar redshifts over the five-year operation. At low redshift  $z \sim 0.2$ , DESI will conduct a magnitude-limited Bright Galaxy Survey (BGS). DESI will measure luminous red galaxies (LRGs) upto redshift  $z < 1$  and the emission line galaxies (ELG) will be one of the DESI targets with redshift upto  $z < 1.7$ . Quasars will

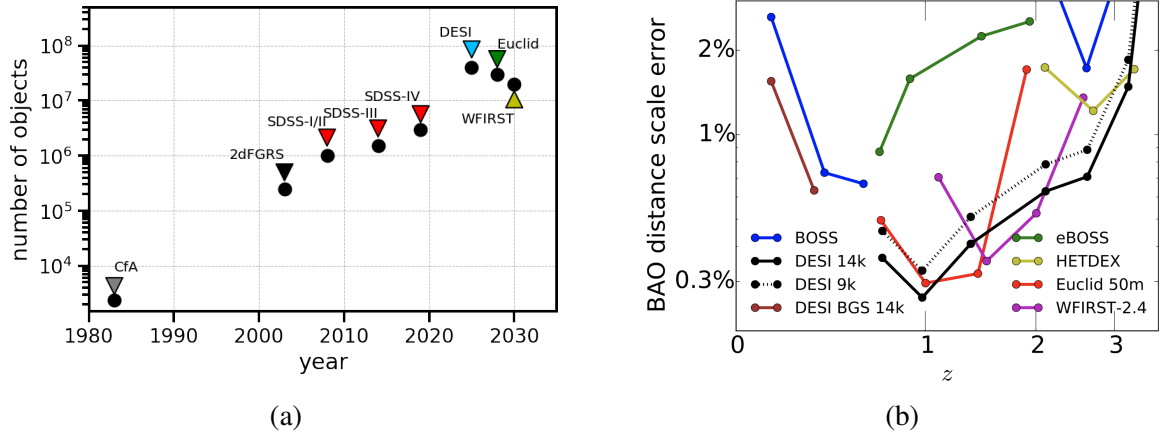


Fig. 1.5 Left panel: number of objects with spectroscopic confirmed redshifts obtained by the galaxy redshift survey over the past decades. Right panel: Error on the BAO scale as a function of redshift presented in comparable bins for DESI, BOSS, Euclid, WFIRST, HETDEX, and eBOSS (Image credit: (DESI Collaboration et al., 2016a)).

also be targeted by DESI as direct tracers of the underlying matter field and as tracers of the neutral hydrogen distribution by  $\text{Ly}\alpha$  forest absorption. Compared to Euclid, DESI is more powerful in constraining cosmology at redshift  $z < 1$  and at redshift  $z > 2$  with the  $\text{Ly}\alpha$  objects, while EUCLID is more powerful at redshift  $1 < z < 2$  with its focus on  $\text{H}\alpha$  emitting galaxies. The DESI experiment is also complementary to the imaging surveys such as Stage-III Dark Energy Survey (DES, operating 2013-2018) and the Stage-IV Large Synoptic Survey Telescope (LSST, scheduled early next decade; Collaboration et al., 2009). Right panel in Fig. 1.4 shows DESI forecast on the linear growth rate as a function of redshift. The plot demonstrates that there will be realistic possibility for the future surveys to distinguish certain modified gravity scenarios from the current standard cosmology model.

Apart from the origin of the cosmic expansion, the constraint on the total neutrino mass will also be addressed by the future galaxy redshift surveys. Combined with the Planck CMB data, DESI is expected to constrain the sum of neutrino mass with a resolution of 0.020 eV. This will be helpful in determining the neutrino hierarchy to obtain the absolute mass scale. The outcome will be a complement to the neutrino mixing experiments such as MINOS Collaboration et al. (2014) and NOvA Collaboration et al. (2017).

The left panel of Fig. 1.5 shows the increasing observed objects with confirmed spectroscopic redshifts over the decades. The right panel of Fig. 1.5 shows the fractional error in the BAO distance measurements at different redshifts for the past and current galaxy surveys (BOSS and eBOSS) and the forecasts for the future surveys (DESI, HETDEX, Euclid, WFIRST (Spergel et al., 2013)).



## 1.5 Outline of the thesis

Chapter 2 briefly discuss the basic theories in cosmology. We will stay in the linear regime and introduce the important concepts in understanding the large scale structure. Chapter 3 extends the discussion to nonlinear regime, there we will introduce the modelling for the two-point statistics, that includes the perturbation theories for the gravitational evolution of the matter field, the bias expansion for the relation between the underlying matter field and the luminous tracers and the impact of the the peculiar velocity of the luminous tracers on modelling the two-point statistics in the nonlinear regime. Chapter 4 begins with an overview of the eBOSS quasar sample, then followed by the methodology. We show various validation tests on the theoretical modelling as well as the observational systematics. Finally, we also discuss the BAO-only fits as a robustness test of the full-shape analysis. Chapter 5 shows the cosmological implications inferred from the eBOSS quasar sample. We summarize our findings in the BAO distance measurements and the cosmic growth rate measurements and focus on the results obtained using DR16 QSO sample. Chapter 6 extends our analysis and show how to RSD as a test of the alternative cosmological models. Chapter 7 summarizes the results and provide future perspectives. Appendix A shows the BAO reconstruction on the QSO catalogue. Appendix B provides supplementary material to the eBOSS analyses.



# Chapter 2

## Theory basics

In this chapter, we begin with a quick review of the basic equations that describes the dynamics of the universe as a smooth background. Then we review the evolution of the density perturbation on top of the smooth background, where we will be mostly interested in the scalar component, as it is most responsible for the structure growth. In the next, we discuss the density perturbation in the early universe in the linear regime, and introduce the concept of the N-point statistics. Finally, a summary on the important concepts for understanding the observational universe focused on clustering analysis will be given.

### 2.1 Einstein equations and general relativity

We briefly review the Einstein field equations, because on the cosmological scale, gravity is supposed to be the force that is supposed to be most relevant for the dynamics of the large scale structure. The Einstein field equation relates the geometry of spacetime to the distribution of matter described by the energy momentum tensor  $T_{\mu\nu}$ ,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (2.1)$$

where  $\Lambda$  is the cosmological constant and the Newton's constant  $G$ . The Einstein tensor is defined as,  $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R}$ , and the Ricci tensor can be expressed using the Christoffel symbols,

$$R_{\mu\nu} = \Gamma_{\mu\nu,\alpha}^{\alpha} - \Gamma_{\mu\alpha,\nu}^{\alpha} + \Gamma_{\beta\alpha}^{\alpha} \Gamma_{\nu\mu}^{\beta} - \Gamma_{\beta\nu}^{\alpha} \Gamma_{\alpha\mu}^{\beta}, \quad (2.2)$$

where the Ricci scalar in the Einstein tensor is the contraction of the Ricci tensor,  $\mathcal{R} \equiv g^{\mu\nu} R_{\mu\nu}$ , and the Christoffel symbol is defined as  $\Gamma_{\mu\nu}^{\sigma} = \frac{1}{2}g^{\sigma\rho} (\partial_{\mu}g_{\nu\rho} + \partial_{\nu}g_{\rho\mu} - \partial_{\rho}g_{\mu\nu})$ .

The cosmological principle states that the universe is homogeneous and isotropic on large scales, the most spatially symmetric metric tensor  $g_{\mu\nu}$  for an expanding universe is the

Friedmann-Robertson-Walker (FRW; [Friedmann, 1924](#); [Robertson, 1935](#); [Walker, 1937](#)) metric. Expressing the FRW metric in spherical coordinate we have

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right], \quad (2.3)$$

with  $a(t)$  being the scale factor. The curvature constant  $K$  characterizes the spatial geometry of the universe and can take the value of  $\{0, -1, 1\}$  for a Euclidean (flat), Hyperbolic (open) or Elliptical (closed) spatial hypersurface of the universe. The idealized distribution of matter (no shear stress, viscosity, heat conduction) can be approximated as a perfect fluid and characterized completely by its pressure and rest frame energy density as

$$T_{\mu\nu} = (p + \rho)U_\mu U_\nu + pg_{\mu\nu}. \quad (2.4)$$

In the particle rest-frame, the components of the four-velocity  $U_\mu = dx_\mu/d\tau$  takes the value  $\{1, 0, 0, 0\}$ .  $p$  and  $\rho$  are pressure and energy density, respectively. As a result of the energy-momentum conservation equation,  $\nabla^\nu T_{\mu\nu} = 0$ , the continuity equation can be derived as

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \quad (2.5)$$

The continuity equation implies that the energy density follows  $\rho \propto a^{-3(1+w)}$ , with the equation of state parameter  $w \equiv \frac{p}{\rho}$ .

Applying the FRW metric defined in Eq. (2.3), we find the non-vanishing components of the Einstein tensor to be  $G_0^0$  and  $G_j^j$ . Together with the assumption of the perfect fluid, this leads to the two Friedmann equations as

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{K}{a^2}, \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p). \end{aligned} \quad (2.6)$$

It is convenient to define the Hubble parameter  $H \equiv \frac{\dot{a}}{a}$ . The set of Friedmann equations describe the components with different equations of state parameters and consequently they have different dependence on the scale factor, for radiation  $w_\gamma = 1/3$  and  $\rho_\gamma \propto a^{-4}$ , for matter  $w_m = 0$  and  $\rho_m \propto a^{-3}$ . To explain the accelerated expansion of the Universe, we need an energy component with  $\rho + 3p \leq -1$  (see Eq. 2.6) and the simplest solution is to assume a constant energy density over the cosmic time with  $w_{de} = -1$ . It is also useful to define a density parameter,

$$\Omega_i \equiv \frac{\rho_{i,0}}{\rho_{\text{crit}}} = \frac{8\pi G}{3H_0^2} \rho_{i,0}, \quad \Omega_K = -\frac{K}{H_0^2} \quad (2.7)$$

where  $\rho_{\text{crit}}$  is the critical density, which is the density necessary so the expansion rate of the Universe is sufficient to prevent a collapse. The critical density is given by

$$\rho_{\text{crit}} \equiv \frac{3H_0^2}{8\pi G} = 2.773 \times 10^{11} \text{M}_\odot h^{-1} (\text{Mpc} h^{-1})^{-3} \sim 6 \text{ protons} / \text{m}^3. \quad (2.8)$$

### 2.1.1 Inhomogeneity in a smooth background

So far we have assumed that the universe is homogeneous. In reality, the constituents of the universe are not uniformly distributed. The quantum vacuum oscillation on microscopic scales is believed to be the origin of the primordial inhomogeneity. Also, the spacetime is not guaranteed to be flat. In the case where the deviation of the spacetime from FRW form is small, we can introduce small space- and time- dependent metric perturbation  $\delta g_{\mu\nu}(\tau, x)$  to the un-perturbed homogeneous background  $\bar{g}_{\mu\nu}$  as,

$$ds^2 = [\bar{g}_{\mu\nu} + \delta g_{\mu\nu}(\tau, x)] dx^\mu dx^\nu. \quad (2.9)$$

We know that  $g_{\mu\nu}$  has 10 degrees of freedom (d.o.f.). The components of the metric can be further decomposed using the scalar-vector-tensor decomposition (SVT) by identifying their behaviours under spatial rotations. We can find four scalar components with 1 d.o.f. for each, two vector components with 2 d.o.f. for each, and one tensor component with 2 d.o.f.. Each type of the perturbations after the SVT decomposition evolves independently. Since the rotational velocity field is usually set to be zero, the vector mode is not excited. In the following we are only going to focus on the scalar and the tensor component.

#### Evolution of scalar perturbation

The scalar perturbations are particularly interesting, because they are the seed of the density field of the large scale structure. We firstly focus on the scalar modes and introduce a scalar field with a small perturbation  $\varphi(\mathbf{x}, t) = \bar{\varphi}(t) + \delta\varphi(\mathbf{x}, t)$ . The scalar part of the perturbed metric reads as

$$g_{\mu\nu} = a^2 \begin{pmatrix} -(1+2\phi) & \partial_i B \\ \partial_i B & (1-2\psi)\delta_{ij} + D_{ij}E \end{pmatrix}. \quad (2.10)$$

The perturbation in the field is tightly coupled to the perturbation in the metric as well as the energy momentum tensor, which in the case of a scalar field can be derived from Lagrangian,

$$T_{\mu\nu} = \frac{\partial \mathcal{L}_\varphi}{\partial g^{\mu\nu}} + g_{\mu\nu} \mathcal{L}_\varphi, \quad (2.11)$$

where the Lagrangian with a potential  $V(\varphi)$  reads as

$$\mathcal{L}_\varphi = -\frac{1}{2} g^{\mu\nu} \frac{\partial \varphi}{\partial x^\mu} \frac{\partial \varphi}{\partial x^\nu} - V(\varphi). \quad (2.12)$$

Combine Eq. (2.11) and (2.12), the energy momentum tensor for a canonical scalar field is given by,

$$T_\beta^\alpha = g^{\alpha\nu} \frac{\partial \varphi}{\partial x^\nu} \frac{\partial \varphi}{\partial x^\beta} - g_\beta^\alpha \left[ \frac{1}{2} g^{\mu\nu} \frac{\partial \varphi}{\partial x^\mu} \frac{\partial \varphi}{\partial x^\nu} + V(\varphi) \right]. \quad (2.13)$$

Follow [Riotto \(2002\)](#) (chapter 7) and [Sánchez \(2018\)](#) (chapter 2) we can find that the components of the perturbed energy momentum tensor  $\delta T_\nu^\mu = \delta g^{\mu\alpha} T_{\alpha\nu} + g^{\mu\alpha} \delta T_{\alpha\nu}$  are given by ([Riotto, 2002](#)),

$$\begin{aligned} \delta T_0^0 &= \phi \bar{\varphi}'^2 - \delta \varphi' \bar{\varphi}' - \delta \varphi \frac{\partial V}{\partial \varphi} a^2 \\ \delta T_0^i &= \partial^i B \bar{\varphi}'^2 + \partial^i \delta \varphi \bar{\varphi}' \\ \delta T_i^0 &= -\partial^i \delta \varphi \bar{\varphi}' \\ \delta T_j^i &= \left( -\phi \bar{\varphi}'^2 + \delta \varphi' \bar{\varphi}' - \delta \varphi \frac{\partial V}{\partial \varphi} a^2 \right) \delta_j^i. \end{aligned} \quad (2.14)$$

The perturbations in the metric and energy-momentum tensors are related via the Einstein field equations for the linearised perturbed parts  $\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu}$ . In the case of the scalar perturbation, there is essentially only 2 physical d.o.f (each of the 4 scalar component is coordinate dependent, the gauge choice of the time or the spatial coordinate can fix 2 of the 4 scalar components), we are allowed to choose the coordinate (gauge fixing) so that we can uniquely identify the scalar coordinate transformations. In the following we always prefer to work in the longitudinal gauge (also called conformal-Newtonian Gauge). This gauge is most intuitive to the observers who are attached to the unperturbed frame and is especially suitable to study perturbations in the sub-horizon scales. The longitudinal gauge implies  $B = E = 0$  and allows us to equalize the metric perturbation to the gauge invariant variables, the so called Bardeen potentials  $\Phi = \phi$ ,  $\Psi = \psi$ . Assume the energy momentum tensor to be diagonal (no stress,  $\Psi = \Phi$ ), we arrive at the following set of equations derived from the linearised Einstein equations,

$$\nabla^2 \Phi - 3\mathcal{H} (\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \delta T_0^0, \quad (2.15)$$

$$\nabla_i (\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \delta T_i^0, \quad (2.16)$$

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi G a^2 \delta T_{ij}, \quad (2.17)$$

with  $\mathcal{H} = aH$  being the comoving Hubble parameter. Fourier transform Eq.(2.15)-Eq.(2.17) and use Eq. (2.14), the evolution of the potential in Fourier space reads

$$\Phi_{\mathbf{k}}'' + 2\left(\mathcal{H} - \frac{\phi''}{\phi'}\right)\Phi_{\mathbf{k}}' + 2\left(\mathcal{H}' - \mathcal{H}\frac{\phi''}{\phi'}\right)\Phi_{\mathbf{k}} + k^2\Phi_{\mathbf{k}} = 0. \quad (2.18)$$

Up till this point, we are only talking about a generic massless scalar field. In the next, I would like to be more specific and see what if the dynamics happens under the context of the inflationary phase, i.e., this massless scalar field is precisely the inflaton field. During the inflation phase, the space undergoes rapid expansion, the comoving Hubble radius significantly decreases compared to the wavelength of the perturbation mode by mode. Once the mode exits the horizon, it "freezes" and the solution to Eq. (2.18) on super-horizon scale is  $\Phi_{\mathbf{k}} \propto H \frac{\delta\varphi_{\mathbf{k}}}{\dot{\phi}}$ . Now it becomes interesting to find the evolution of the scalar field in Fourier space, which can be derived from the perturbed Klein-Gordon equation as

$$\ddot{\delta\varphi}_{\mathbf{k}} + 3H\dot{\delta\varphi}_{\mathbf{k}} + \frac{k^2}{a^2}\delta\varphi_{\mathbf{k}} + \partial_{\varphi}^2 V \delta\varphi_{\mathbf{k}} = 0, \quad (2.19)$$

with the solution on super-horizon scale  $|\delta\varphi_{\mathbf{k}}| \simeq \frac{H}{\sqrt{2k^3}} \Big|_{k=aH}$ . So far we restrict ourselves within the classical analysis of the fluctuation and this gives a very good description how the initial perturbation are stretched into large scales. However, as we seek for the origin of the density perturbation, it requires further to quantize the cosmological perturbation, and an important task is to find an action written in terms of gauge invariant variables. As shown in [Brandenberger et al. \(1993\)](#); [Mukhanov et al. \(1992\)](#), the perturbation leads to an additional quadratic correction to the background action<sup>1</sup>

$$\delta^{(2)}S = \frac{1}{2} \int d^4x \left\{ v'^2 - v_{,i}v_{,j}\delta^{ij} + \frac{\mathcal{Z}''}{\mathcal{Z}}v^2 \right\}, \quad (2.20)$$

where  $\mathcal{Z} = a \frac{\bar{\varphi}'}{\mathcal{H}}$ . The combination of metric and field fluctuation  $v = a \left( \frac{\bar{\varphi}'}{\mathcal{H}}\psi + \delta\varphi \right) = \mathcal{Z}\mathcal{R}$  is exactly what we look for, with  $\mathcal{R}$  being the comoving curvature perturbation. Promoting the classical field to an operator with creation  $\hat{a}_k^\dagger$  and annihilation operators  $\hat{a}_k$ ,

$$\hat{v} = \frac{1}{2} \frac{1}{(2\pi)^{3/2}} \int d^3k \left[ e^{ikx} v_k^*(\eta) \hat{a}_k + e^{-ikx} v_k(\eta) \hat{a}_k^\dagger \right]. \quad (2.21)$$

<sup>1</sup>this correction can be derived classically as well.

The equation of motion for the mode function  $v_k$  field can be obtained by making use of the variational principle w.r.t.  $v$  in Eq. (2.20),

$$v_k'' + \left( k^2 - \frac{\mathcal{Z}''}{\mathcal{Z}} \right) v_k = 0. \quad (2.22)$$

To solve Eq. (2.22) is rather involved, I will skip the derivation but directly list the solution on super-horizon scales. Express the solution in terms of the comoving curvature perturbation evaluated at the horizon exit for a given mode read

$$\mathcal{R} = \frac{H}{\dot{\phi}} \frac{1}{\sqrt{2k^3}} \Big|_{k=aH}. \quad (2.23)$$

The full dimensionless power spectrum reads,

$$\Delta_{\mathcal{R}}^2(k) = \frac{k^3}{2\pi^2} |\mathcal{R}|^2 = \frac{H^2}{\dot{\phi}^2} \left( \frac{H}{2\pi} \right)^2 \left( \frac{k}{aH} \right)^{n_s-1}, \quad (2.24)$$

where  $n_s$  is the scalar spectral index. This is one of the most important quantity that links the early universe after the inflation to the late time observable universe.

### Evolution of the tensor modes

The tensor modes are related to the gravitational waves. The procedure for deriving the dynamics of the tensor modes is in analogy to the derivations for the scalar modes. Ever since the first detection by the LIGO collaboration ([Abbott et al., 2016a](#)), the gravitational wave detection opened a new frequency window and has several important cosmological implications. The multi-messenger observations using the gravitational wave in combination with the electromagnetic signals serves as a standard siren and provides the constraints on the local measurement of the Hubble parameter. The gravitational wave from the binary merger event can be to test the general relativity and the speed of the gravitational wave sets limits on the mass of the gravitons ([Abbott et al., 2016b](#)) that can be used to constrain alternative theories than general relativity.



## 2.2 Density perturbation in the early time – linear regime

### 2.2.1 Equations of motion

The scalar fluctuation during the inflation is related to the density perturbation after the inflation. The conservation of the energy momentum tensor leads to the general relativistic version of the continuity and Euler equations. We focus on the matter and the radiation, the continuity and the Euler equation in Fourier space takes the form as

$$\delta'_m - ikv_m = -3\Phi', \quad (2.25)$$

$$\frac{3}{4}\delta'_\gamma + ikv_\gamma = 3\Phi', \quad (2.26)$$

$$v'_m + \mathcal{H}v_m = -ik\Phi, \quad (2.27)$$

$$v'_\gamma + \frac{1}{4}\delta'_\gamma = -\Phi. \quad (2.28)$$

The evolution of these equation can be solved in the limit of the super-horizon and sub-horizon. For the super-horizon mode  $k \ll \mathcal{H}$ , the  $k$ -dependence can be neglected and the potential  $\Phi$  is almost a constant except for a factor 9/10 drop at the transition between radiation-dominated to matter-dominated era. The density fluctuation is related to the potential by a integration constant in the longitudinal gauge<sup>2</sup>. For the sub-horizon modes  $k \gg \mathcal{H}$ , we need to distinguish the evolution between the radiation-dominated and matter-dominated era. During the radiation-dominated era the potential decays and then oscillate while the matter perturbation still grows logarithmically as the conformal time. During the matter-dominated era, the potential on the sub-horizon scale is a constant as long as we ignore the contribution of the dark energy (which cause the potential to decay)  $\Phi' = 0$ , combining Eq. (2.25) and (2.27) one can arrive at an equation that describe the dynamics of the structure growth,

$$\frac{d^2}{d\tau^2}D(\tau) + \mathcal{H}\frac{d}{d\tau}D(\tau) - \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2D(\tau) = 0, \quad (2.29)$$

where the solution is the time dependence growth factor  $D(a)$ , that is important to understand the structure growth under the linear approximation,

$$D(a) = \frac{5\Omega_m}{2} \frac{H(a)}{H_0} \int_0^a \frac{da_1}{(a_1 H(a_1)/H_0)^3}, \quad (2.30)$$

---

<sup>2</sup>density fluctuation depends on th gauge choice

### 2.2.2 N-point statistics

To quantify the distribution of large scale structure, one way to compress the information is the N-point statistics. Starting from the 2-point statistics, with the probability of finding a pair of points in volume elements  $dV_{1,2}$  separated by a distance  $r_{12}$  being,

$$\begin{aligned} dP_{12} &= \langle \rho_1 \rho_2 \rangle dV_1 dV_2 \\ &= \bar{\rho}^2 (1 + \xi(r_{12})) dV_1 dV_2, \end{aligned} \quad (2.31)$$

where the two point correlation function  $\xi(r_{12})$  denotes the probability excess of finding a pair for the given separation distance  $r_{12}$ ,

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (2.32)$$

where the density contrast is defined as the spatial density over the mean background  $\delta(x) \equiv (\rho(x) - \bar{\rho})/\bar{\rho}$ . The analogue in Fourier space is the power spectrum:

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P(k). \quad (2.33)$$

As we define the Fourier transform as,  $\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k \delta(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}}$ , the two point correlation function and the power spectrum are related as

$$P(k) = \int \xi(r) e^{i\mathbf{k} \cdot \mathbf{r}} d^3r. \quad (2.34)$$

The definition of the two point correlation function can be easily generalized to n-point statistics,

$$dP_{\{1..n\}} = \bar{\rho}^n (1 + \xi^{(n)}) dV_1 \dots dV_n, \quad (2.35)$$

where the n-point correlation function  $\xi^{(n)}$  can be defined as

$$1 + \xi^{(n)} = \left\langle \prod_{i=1}^n (1 + \delta_i) \right\rangle. \quad (2.36)$$

The higher order statistics can be used to constrain the single field slow-roll inflationary model via detecting the primordial non-Gaussianity (Desjacques & Seljak, 2010). The higher order statistics can also be used to test the topological defect models, which could be formed at phase transition. The topological defects produce distinctive non-Gaussian signatures (Brandenberger et al., 1993) that leaves an imprint on the large scale structures (Verde et al., 2001).

## 2.3 Modelling the observational LSS clustering

Measuring the distances in the spatial coordinate is one of the most intuitive perception in our daily experience. However, this simple quantity becomes not straightforward at the cosmic scale as the coordinate of the universe is not uniquely fixed. On cosmological scales, it is more convenient to introduce the cosmological redshift. The redshift defined as the wavelength emitted by a distant object compared to the one in the rest frame as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}}. \quad (2.37)$$

The ratio between the rest-frame wavelength and the observed wavelength is the scale factor  $a$ , which is related to the redshift as  $a = (1 + z)^{-1}$ . The distance on the cosmological scale is impacted by the expansion history. One of the most commonly used distance is the comoving distance. In the Euclidean Universe, the comoving angular distance is the light travel during the conformal time,

$$D_M = \int_{t_e}^t c \frac{dt'}{a(t')} = \int_0^z \frac{cdz'}{H(z')} \quad (2.38)$$

where the second equality relates the distance to the cosmological redshift, and the Hubble parameter is determined by the density content of various components (see Eq. (2.6)). Hubble distance is also commonly used in cosmology and is defined as

$$D_H = c/H(z), \quad (2.39)$$

with  $c$  being the speed of the light in the vacuum.

### 2.3.1 Baryonic acoustic oscillation as a standard ruler

In the early universe the photons are energetic enough to scatter off the electrons through Compton scattering. The electrons interact with the protons via coulomb interaction. Effectively, the photons are coupled to the baryon fluid<sup>3</sup>. The slight inhomogeneous distribution of the perturbations, originated from the quantum fluctuation, exerts a spatial varying photon pressure and source a spherical wave that pushes the fluid outwards. This phenomenon is in

---

<sup>3</sup>Although electrons are leptons, the mass of the fluid is dominated by protons and it is named as baryon fluid for simplicity.

analogy to an acoustic wave with the sound speed given by

$$c_s = \sqrt{\frac{1}{3(1+R)}}, \quad (2.40)$$

where  $R(z) = 3\Omega_b/4\Omega_r(1+z)$  is the baryon-to-photon ratio. At the epoch of decoupling<sup>4</sup> the photons are no longer bond to the baryon plasma and free stream away, leaving the matter frozen in its place. The scale the sound can travel from the beginning of the universe until the decoupling is the sound horizon  $r_{\text{drag}}$  at drag epoch  $z_{\text{drag}}$ ,

$$r_{\text{drag}}(z_{\text{drag}}) = \int_0^{\tau(z_{\text{drag}})} \frac{d\tau}{\sqrt{3(1+R)}}. \quad (2.41)$$

The sound horizon is a fixed scale and can be precisely measured using the CMB data, therefore, the BAO can be used as a standard ruler to determine the cosmic distances.

Given the isotropic feature of the BAO, one can perform an Alcock-Paczynski (AP) test (Alcock & Paczynski, 1979) to determine the Hubble parameter  $H(z)$  and the angular diameter distance  $D_A(z)$ . As we assume a spherical object in the sky with fixed physical diameter  $\delta l$ , the angular size is given by

$$\delta\theta = \frac{L_{\text{AB}}}{D_A(z)} = \frac{1}{1+z} \frac{r_s}{D_A(z)}. \quad (2.42)$$

On the other hand, looking through the object the redshift difference of the object between the two ends is given by  $\delta z = z_B - z_A = \frac{a_0}{a(t_B)} - \frac{a_0}{a(t_A)} \approx a_0 \frac{\dot{a}(\bar{t})}{a^2(\bar{t})} \Delta t_{\text{AB}} = \frac{H(\bar{z})}{c} \frac{a_0 L_{\text{AB}}}{a(\bar{z})}$ . Express the redshift difference in comoving coordinate with the sound horizon we have

$$\delta z = \frac{r_s H(z)}{c}. \quad (2.43)$$

If we have absolute no access to the intrinsic size of the object  $L_{\text{AB}}$ , combine Eq. (2.42) and Eq. (2.43) we can obtain the product of the angular diameter distance and the Hubble parameter as

$$F_{\text{AP}}(z) \equiv \frac{1+z}{c} D_A(z) H(z) = \frac{\delta z}{\delta\theta}. \quad (2.44)$$

As pointed out in the previous paragraph that, the sound horizon  $r_{\text{drag}}$  scale can be calculated for a given baryon-to-photon ratio. With this information, we can break the degeneracy between the the Hubble parameter  $H(z)$  and the comoving angular diameter distance  $D_M(z)$ .

<sup>4</sup>In literature this is also called recombination, which is confusing enough since this the first time electron combined with proton

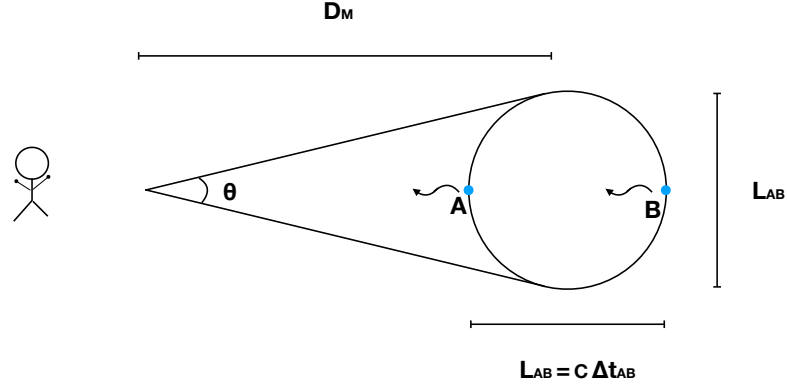


Fig. 2.1 A schematic drawing of the AP test.

### 2.3.2 RSD in linear regime

When we observe the galaxies in the redshift space, the radial distances of the galaxies is affected by the peculiar velocity induced by the gravitational attraction from the surrounding environment of the galaxies. The observed wavelength  $\lambda_{\text{obs}}$  receive an extra component apart from the stretching due to the cosmic expansion and we have  $\lambda_{\text{obs}} = (1 + \frac{v_{\parallel}}{c})\lambda_{\text{cos}}$ , where  $v_{\parallel}$  is the component parallel along the line of sight direction. Consequently, the observed redshift is modified to be  $1 + z_{\text{obs}} = (1 + z_{\text{cos}})(1 + \frac{v_{\parallel}}{c})$ . Since the extra component has an impact only in the radial direction, the clustering amplitude becomes an angle-dependent quantity, which is one of the most important reason for the anisotropic clustering. The strength in the distortion is controlled by the growth rate for structure formation. In this way, the redshift distortion provides unique way to measure the growth rate. During the matter dominated epoch, the potential  $\Phi$  appear on the right hand side of Eq. (2.25) is a constant on the sub-horizon scales and the matter continuity equation reduces to

$$\delta'_m + ikv_m = 0. \quad (2.45)$$

Using  $d/d\tau = a^2 H d/da$ , the velocity field can be related to the density fluctuation field as

$$v_m(k, a) = ik \frac{f(a)aH}{k^2} \delta_m(k, a), \quad (2.46)$$

where the linear growth rate is given as the derivative of growth factor w.r.t the scale factor (see Eq. (2.30)) on logarithmic scale,

$$f \equiv \frac{d \ln D(a)}{d \ln a}. \quad (2.47)$$

Given the number conservation in the redshift space coordinate  $d\mathbf{s}$  and real space coordinate  $d\mathbf{r}$ ,  $\rho_m^s(\mathbf{s})d^3s = \rho_m(\mathbf{r})d^3r$ , the Jacobian that map between the redshift and the real space density fluctuation field  $(1 + \delta^s(\mathbf{s}))d^3s = (1 + \delta(\mathbf{r}))d^3r$  is given by

$$J = \left| \frac{d\mathbf{r}}{d\mathbf{s}} \right| = \left| \frac{r^2 dr}{s^2 ds} \right| = \left\{ 1 + \frac{1 + z_{\cos}}{H(z_{\cos})} \frac{v_{\parallel}}{r} \right\}^{-2} \left\{ 1 + \frac{1 + z_{\cos}}{H(z_{\cos})} \frac{\partial v_{\parallel}}{\partial r} \right\}^{-1}. \quad (2.48)$$

The displacement of the galaxies  $(1 + z_{\cos})v_{\parallel}/H(z_{\cos})$  (typically at the order of  $< \sim 10 h^{-1} \text{Mpc}$ ) in the first term is typically much smaller compared to the radial distance of the galaxies at  $r \approx D_M(z_{\text{eff}})$  towards us, where  $z_{\text{eff}}$  is the effective redshift for a galaxy sample. Taking eBOSS LSS quasar for an example, the effective redshift  $z_{\text{eff}} \sim 1.5$  corresponds to a comoving angular diameter distance of  $\sim 3000 h^{-1} \text{Mpc}$ . Therefore, we ignore the first term can approximate the Jacobian as

$$J \approx \left\{ 1 + \frac{1 + z_{\cos}}{H(z_{\cos})} \frac{\partial v_{\parallel}}{\partial r} \right\}^{-1}. \quad (2.49)$$

We can link the density field with the velocity field using the relation in Eq. (2.46), in Fourier space with  $\partial/\partial r_i \Rightarrow i k_i$ , the overdensity in the redshift space can be related to the real space as

$$\delta^s(\mathbf{k}) = \delta(\mathbf{k}) \left( 1 + f(z) \mu_k^2 \right), \quad (2.50)$$

where  $\mu_k = k_{\parallel}/k$  and because the density field and the velocity field is related by the linear growth rate,  $f(z)$  also enters the mapping between the real and redshift space coordinate. The linear growth rate can be very well parameterized by the matter density parameter,

$$f(z) = [\Omega_m(z)]^\gamma, \quad (2.51)$$

the exponent  $\gamma$  can be derived by approximating the presence of only matter and dark energy. Follow [Linder & Cahn \(2007\)](#) and assuming general relativity, we have the following relation

$$\gamma \simeq \frac{6 - 3(1 + w_{\text{DE}})}{11 - 6(1 + w_{\text{DE}})} \simeq \frac{6}{11} + \frac{3}{121} (1 + w_{\text{DE}}). \quad (2.52)$$

For  $\Lambda$ CDM cosmology with constant equation of state parameter  $w_{\text{DE}} = -1$ , the dependence is reduced to  $f(z) = [\Omega_{\text{M}}(z)]^{0.55}$ . Since the linear growth rate is tightly connected to the redshift space distortion measurement, the anisotropic clustering induced by RSD can be used to test the deviation from the GR prediction (more detail see Chapter 6).





## Chapter 3

# Modelling the 2-point correlation function

The modelling of the two point statistics includes three ingredients, the nonlinear gravitational evolution, the bias expansion that establishes the relation between the luminous tracers and the underlying matter field, and the nonlinear modelling of the RSD. The model is constructed in the Fourier space and Fourier transformed to obtain the two-point correlation function. The full model in redshift space can be expressed as

$$P^s(k, \mu) = F_{\text{FOG}}(k, \mu) \exp \left[ - (k \mu \sigma_{\text{zerr}})^2 \right] P_{\text{novir}}(k, \mu), \quad (3.1)$$

where the first term  $F_{\text{FOG}}(k, \mu)$  is the finger-of-god factor that describes the redshift space distortion on the small scales (see Section 3.3.2).  $P_{\text{novir}}$  is a model for the redshift space power spectrum on the large scales (see Section 3.3.1). To obtain this term, we need a bias expansion of the tracer density field (Section 3.2), we also need to input the nonlinear power spectrum for the matter and the velocity field. In Section 3.1.1–3.1.3 we will review the perturbation theories that were used to predict the nonlinear gravitational evolution. For the analysis of the final eBOSS quasar sample, we apply a new hybrid approach to calculate the power spectrum in redshift space, where the nonlinear matter power spectrum is described in Section 3.1.4 and the velocity power spectrum is discussed in Section 3.1.5. Lastly, we introduce a Gaussian damping term with a scale-independent free parameter  $\sigma_{\text{zerr}}$  to account for the redshift uncertainty for the quasar sample.

## 3.1 Gravitational evolution in the nonlinear regime

### 3.1.1 Standard perturbation theory

The evolution of the perturbations on the large scales is dominated by the cold dark matter. Assuming the cold dark matter matter to be collision-less fluid, its dynamics is governed by the Vlasov equation as the Boltzmann equation in the collision-less limit. Taking the moments of the Vlasov equation, one can arrive at the continuity equation and the Euler equation. Combine these two equations with the Poisson equation and assume there is no shell-crossing (single-streaming approximation), the set of equations can be written as

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \nabla \cdot \{(1 + \delta(\mathbf{x}, \tau))\mathbf{v}(\mathbf{x}, \tau)\} = 0 \quad (3.2)$$

$$\frac{\partial \mathbf{v}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H}(\tau)\mathbf{v}(\mathbf{x}, \tau) + \mathbf{u}(\mathbf{x}, \tau) \cdot \nabla \mathbf{v}(\mathbf{x}, \tau) = -\nabla \phi(\mathbf{x}, \tau) - \frac{1}{\rho} \nabla_j (\rho \sigma_{ij}) \quad (3.3)$$

$$\nabla^2 \phi(\mathbf{x}, \tau) - \frac{3}{2} \Omega_M(\tau) \mathcal{H}^2(\tau) \delta(\mathbf{x}, \tau) = 0. \quad (3.4)$$

For pressure-less perfect fluid, stress tensor contribution  $\sigma_{ij}$  can be sent to zero. As we write the set of equations in Fourier space and introduce the velocity divergence field  $\theta \equiv \nabla \cdot \mathbf{v}(\mathbf{x}, \tau)$ , we can obtain the coupled equation as follow (Bernardeau et al., 2002),

$$\frac{\partial \tilde{\delta}(\mathbf{k}, \tau)}{\partial \tau} + \tilde{\theta}(\mathbf{k}, \tau) = - \int d^3 k_1 d^3 k_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \alpha(\mathbf{k}_1, \mathbf{k}_2) \tilde{\theta}(\mathbf{k}_1, \tau) \tilde{\delta}(\mathbf{k}_2, \tau) \quad (3.5)$$

$$\begin{aligned} \frac{\partial \tilde{\theta}(\mathbf{k}, \tau)}{\partial \tau} + \mathcal{H}(\tau) \tilde{\theta}(\mathbf{k}, \tau) + \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \tilde{\delta}(\mathbf{k}, \tau) = \\ - \int d^3 k_1 d^3 k_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \beta(\mathbf{k}_1, \mathbf{k}_2) \tilde{\theta}(\mathbf{k}_1, \tau) \tilde{\theta}(\mathbf{k}_2, \tau), \end{aligned} \quad (3.6)$$

where

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_{12} \cdot \mathbf{k}_1}{k_1^2}, \quad \beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{k_{12}^2 (\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2} \quad (3.7)$$

Eq. (3.5) and (3.6) can be expanded perturbatively as,

$$\delta(\mathbf{k}, \tau) = \sum_{n=1}^{\infty} \delta^{(n)}(\mathbf{k}, \tau), \quad \theta(\mathbf{k}, \tau) = \sum_{n=1}^{\infty} \theta^{(n)}(\mathbf{k}, \tau). \quad (3.8)$$

One can write the ansatz for Eq. (3.5) and (3.6) as

$$\begin{aligned}\delta^{(n)}(\mathbf{k}, \tau) &= \int_{\mathbf{k}_1} \cdots \int_{\mathbf{k}_n} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_{12\dots n}) F_n(\mathbf{k}_1, \dots, \mathbf{k}_n, \tau) \delta^{(1)}(\mathbf{k}_1, \tau) \cdots \delta^{(1)}(\mathbf{k}_n, \tau), \\ \theta^{(n)}(\mathbf{k}, \tau) &= -\mathcal{H}(\tau) f(\tau) \int_{\mathbf{k}_1} \cdots \int_{\mathbf{k}_n} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_{12\dots n}) G_n(\mathbf{k}_1, \dots, \mathbf{k}_n, \tau) \delta^{(1)}(\mathbf{k}_1, \tau) \cdots \delta^{(1)}(\mathbf{k}_n, \tau).\end{aligned}\quad (3.9)$$

At the first order, the right hand side of Eq. (3.5) and (3.6) can be set to zero and it is easy to see that  $F_n = G_n = 1$ . At second order, the  $F_n$  and  $G_n$  kernels have the form as

$$\begin{aligned}F_2(\mathbf{k}_1, \mathbf{k}_2) &= \frac{17}{21} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \left[ \left( \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \right)^2 - \frac{1}{3} \right], \\ G_2(\mathbf{k}_1, \mathbf{k}_2) &= \frac{13}{21} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{4}{7} \left[ \left( \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \right)^2 - \frac{1}{3} \right],\end{aligned}\quad (3.10)$$

where the first constant is the monopole contribution that is directly proportional to  $\propto \delta^2$  and corresponds the spherical collapse. The second term is the dipole contribution, it originates from the  $\mathbf{v} \cdot \nabla$  term in the equation of motion and represents the bulk flow, and the third term is the quadrupole contribution that distorts the shape given a gravitational tidal force. This solution is exact in Einstein de Sitter (EdS) Universe, meanwhile it is also good description for  $\Lambda$ CDM even dark energy model with the general time-varying equation of state (Takahashi, 2008). The power spectrum can be consequently expanded as

$$P(k, z) = D_+^2(z) P_0(k) + P_{1\text{-loop}}(k, z) + P_{2\text{-loop}}(k, z) + \dots, \quad (3.11)$$

where  $P_0(k)$  is the initial power spectrum, the loop corrections are the combination of the products in the expansion of the density field. The explicit expression for the next-leading-order term in the case of matter density field can be written as

$$P_{\delta\delta}^{1\text{-loop}}(k) = P_{\delta\delta}^{(22)}(k) + 2P_{\delta\delta}^{(13)}(k), \quad (3.12)$$

where

$$P_{\delta\delta}^{(22)}(k) \equiv \langle \delta^{(2)}(\mathbf{k}) \delta^{(2)}(\mathbf{k}') \rangle = 2 \int_{\mathbf{q}} [F_2(\mathbf{p}, \mathbf{k} - \mathbf{p})]^2 P_0(q) P_0(|\mathbf{k} - \mathbf{q}|) \quad (3.13)$$

$$P_{\delta\delta}^{(13)}(k) \equiv \langle \delta^{(1)}(\mathbf{k}) \delta^{(3)}(\mathbf{k}') \rangle = 3P_0(k) \int_{\mathbf{q}} F_3(\mathbf{q}, -\mathbf{q}, \mathbf{k}) P_0(q). \quad (3.14)$$

Following the same spirit, the similar terms can be written down for the 2-loop corrections.

### 3.1.2 Renormalized perturbation theory

The standard perturbation theory (SPT) introduced in Sec.3.1.1 does not converge fast enough as one approaches the small scales. The amplitude of the fluctuation can become large and the prediction tends to break down at the small scales as the neglected higher orders terms can be important. One way to overcome this problem is to sum up an infinite subset of the contribution to the SPT expansion (Crocce & Scoccimarro, 2006). By rewriting Eq. (3.5) and (3.6) into the matrix formalism, it is easier to see that the structure for the evolved density field can be split into a propagator and a mode-coupling term as

$$P(k, \tau) = G^2(k, \tau)P_0(k) + P_{MC}(k, \tau), \quad (3.15)$$

where  $P_{MC}(k, z)$  is the mode coupling power spectrum. The generalized propagator  $G(k, \tau)$  measures the degree of linear transition from the initial field to the final field and is defined as,

$$\left\langle \frac{\delta \Psi_a(\mathbf{k}, a)}{\delta \phi_b(\mathbf{q})} \right\rangle = \delta_D(\mathbf{k} - \mathbf{q}) G_{ab}(k, a), \quad (3.16)$$

where  $a$  and  $b$  correspond to the density field or the velocity field.  $\Psi_{a,b}(\mathbf{k}, s)$  represents the evolved fields, while  $\phi_{a,b}(\mathbf{q})$  stands for the initial ones. At small scales  $k \gg q$ , the nonlinear effects start to become dominant and the generalized propagator is expected to decay to zero. Follow Crocce et al. (2012) an explicit expression can be derived that has a Gaussian and decay as  $k$  grows,

$$G(k, a) \simeq g_{ab}(a) \exp \left[ -\frac{1}{2} k^2 \sigma_s^2 \right], \text{ for } k \sigma_d \gg 1. \quad (3.17)$$

At large scales  $k \ll q$ , the generalized propagator restores the linear version,

$$g_{ab}(a) = \frac{a}{5} \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} - \frac{a^{-3/2}}{5} \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix}, \quad (3.18)$$

the fluctuation of the displacement field is given by

$$\sigma_s^2 \equiv \frac{(a-1)^2}{3} \int \frac{P_0}{q^2} d^3 \mathbf{q}. \quad (3.19)$$

It is important to properly take the non-linearity into account, as the nonlinear gravitational evolution not only damps the BAO signal, but also shift the BAO peak. As pointed out by Crocce & Scoccimarro (2008), there are two reasons to account for the shift in the BAO peak. Firstly, it is due to the convolution with the propagator term if the BAO peak itself is not

symmetric or it is not sharp enough compared to the width of the propagator. Secondly, it is due to the mode coupling term that involves the SPT kernels  $F_2$  (see Eq. (3.13)). The second dipole term in the kernel (see Eq. (3.10)) encodes the transport of matter by velocity field  $\mathbf{v} \cdot \nabla$ , and this is the origin of a derivative-like correlation function  $\xi'$ . Depending on the sign of the  $\xi'$ , the BAO peak shift left (right) if the sign is positive (negative).

This way of separating the power spectrum into a propagator term and a mode coupling term is known as the renormalized perturbation theory (RPT). It was later on generalize to a multi-point propagator framework such as RegPT (Bernardeau et al., 2012), MPT-breeze (Crocce et al., 2012).

### 3.1.3 Galilean-Invariant RPT

The renormalized perturbation theory (RPT) described in Section 3.1.2 is an innovative way of modelling the power spectrum. However, it breaks the Galilean invariance (Peloso & Pietroni, 2013). However, the resummation of the infrared (IR) modes in RPT is only performed on the propagator and leaves the mode coupling terms untouched, which violates the Galilean invariance (Peloso & Pietroni, 2013) (equivalence principle (Creminelli et al., 2013)). The Galilean invariant RPT (gRPT) is proposed to restore the symmetry and to achieve higher accuracy. gRPT starts with the expression with RegPT and minimise the power spectrum with respect to a boost term for each given mode  $k$ . More detail regarding gRPT can be found in Crocce et al. (prep).

### 3.1.4 RESPRESSO

In order to obtain the nonlinear matter power spectrum, a different approach compared to the pure perturbative calculation is based on the response function. For our analysis on the final eBOSS quasar sample, we tried to integrate Rapid and Efficient SPectrum calculation based on RESponSe functiOn (RESPRESSO) (Nishimichi et al., 2017). RESPRESSO has two steps, firstly a phenomenological expression of the response function is proposed, secondly a reconstruction at the power spectrum is applied. The response function  $K(k, q; z)$  quantifies the variation of the nonlinear power spectrum  $\delta P(k; z)$  of mode  $k$  for a given initial perturbation in the linear power spectrum  $\delta P_{\text{lin}}(q; z)$  of mode  $q$  and is defined as

$$K(k, q; z) = q \frac{\delta P(k; z)}{\delta P_{\text{lin}}(q; z)}. \quad (3.20)$$

Nishimichi et al. (2017) proposed the phenomenological model of the response function based on a calibration with respect to the measurement of 1400 simulations. The model

is in analogy with the expression applied in RegPT (Bernardeau et al., 2012) but has the possibility switching to the SPT prediction,

$$K_{\text{model}}(k, q) = \left[ \left( 1 + \beta_{k,q} + \frac{1}{2} \beta_{k,q}^2 \right) K_{\text{tree}}^{\text{SPT}}(k, q) + \left( 1 + \beta_{k,q} \right) K_{1\text{-loop}}^{\text{SPT}}(k, q) + K_{2\text{-loop}}^{\text{SPT}}(k, q) \right] D(\beta_{k,q}), \quad (3.21)$$

with  $\beta_{k,q} = \alpha_k + \alpha_q$ , where  $\alpha_k = \frac{1}{2} k^2 \sigma_d^2$ . The one-dimensional root mean square of the linear displacement field  $\sigma_d$  is given by,

$$\sigma_d^2 = \int \frac{dq}{6\pi^2} P_{\text{lin}}(q). \quad (3.22)$$

The damping factor  $D(x)$  in Eq. (3.21) is designed to cancel the first order contribution at the low  $q$  limit and forces the model to recover the SPT prediction. This is important because in the expansion of the SPT, each loop is built out of the equal time correlator, e.g. at the 1-loop term the positive  $P_{22}$  cancels with the negative  $P_{13}$ . In the  $q \gtrsim k$  regime, the damping term restores the normalized perturbation theory (RPT; Crocce & Scoccimarro, 2006).

$$D(x) = \begin{cases} \exp(-x), & \text{if } K_{\text{model}}(k, q) > 0 \\ \frac{1}{1+x}, & \text{if } K_{\text{model}}(k, q) < 0 \end{cases}, \quad (3.23)$$

The nonlinear power spectrum at a different redshift with different cosmological parameter  $\mathbf{p}_1$  can be viewed as an expansion around the parameter  $\mathbf{p}_0$ , where an accurate power spectrum template is known. The reconstruction of the nonlinear power spectrum at the target cosmology reads

$$P_{\text{nl}}(k; \mathbf{p}_1) \approx P_{\text{nl}}(k; \mathbf{p}_0) + \int d \ln q K(k, q) \times [P_{\text{lin}}(q; \mathbf{p}_1) - P_{\text{lin}}(q; \mathbf{p}_0)]. \quad (3.24)$$

To validate the nonlinear matter power spectrum, we compared it to the MINERVA N-body simulation. These set of 300 simulations were run using GADGET (Springel, 2005) with  $1000^3$  dark matter particles with a box size of  $L = 1.5 h^{-1} \text{Gpc}$  and imposing periodic boundary condition. The cosmological parameters for the MINERVA simulation is  $h = 0.695$ ,  $\Omega_m = 0.285$ ,  $\Omega_b h^2 = 0.02224$ ,  $\Omega_\Lambda = 0.715$ ,  $n_s = 0.9632$ ,  $\sigma_8 = 0.828$  and without including neutrinos (more detail regarding the simulation can be found in Grieb et al. (2016)). Fig 3.1 compare the matter power spectrum of the RESPRESSO and gRPT to the measured MINERVA spectrum. We compare the output at two snapshots  $z = \{0.57, 1.0\}$  shown in the upper and lower panel, respectively. The left columns show the matter power spectrum directly measured from the mean of the 100 simulations with the grey band being the 2% error with prediction by RESPRESSO (orange) and gRPT (green). The dashed black

curve shows the linear matter power spectrum as a comparison to the nonlinear theory. The right column is the ratio between the measured MINERVA matter power to the RESPRESSO (orange) and gRPT (green). Again, the grey band is the 2% error for the ratio. We can see that both approaches are very close to each other and can describe the measurement from the N-body simulation very well.

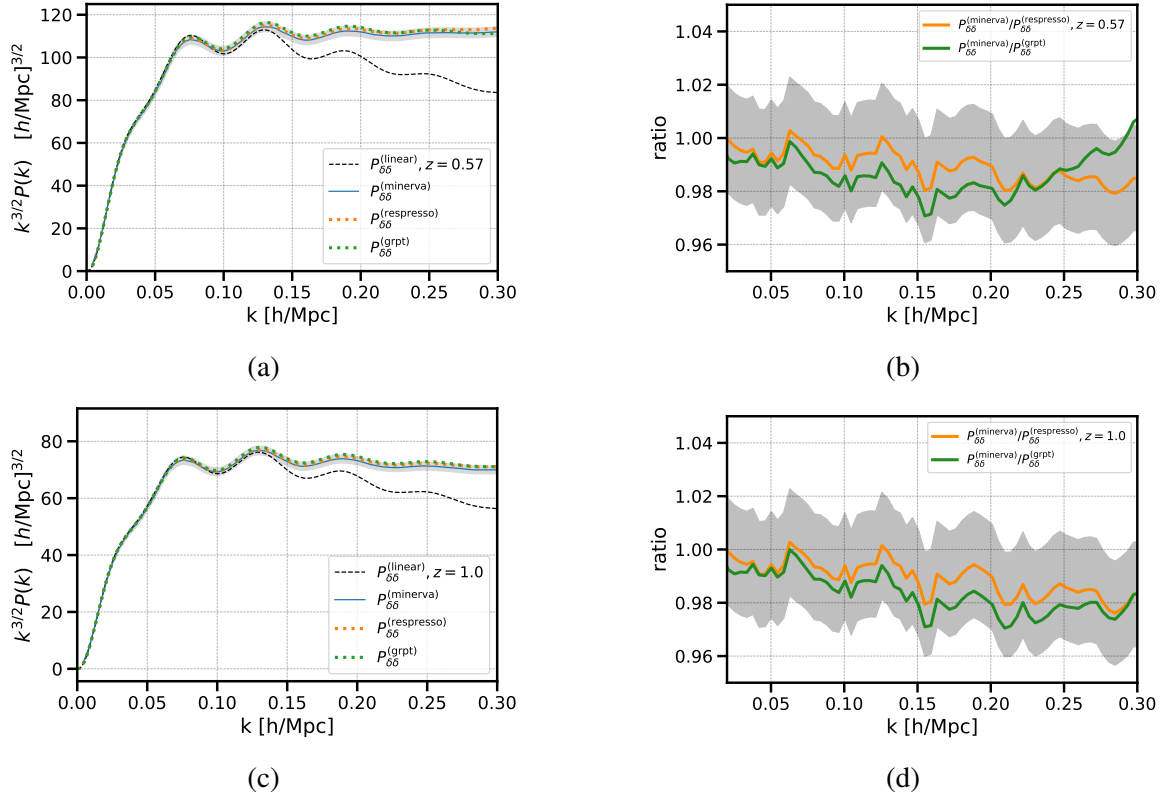


Fig. 3.1 Left: Comparison of the matter power spectrum calculated by the RESPRESSO, gRPT to the measured power from Minerva simulation. The plot also includes a linear matter power as to show the difference between the linear and nonlinear theories. Right: the ratio between the measured power from the simulation and the ones obtained from perturbation theories. Upper panels show the comparison at redshift  $z = 0.57$  while the lower panel shows the comparison at redshift  $z = 1.0$ .

### 3.1.5 Fitting function for the velocity

The full modelling of the power spectrum in the redshift space requires the input of the cross-power spectrum for the matter and velocity divergence  $P_{\delta\theta}$  as well as the auto-power spectrum for the velocity divergence field  $P_{\theta\theta}$ , which is not provided by RESPRESSO at the moment. An alternative approach to obtain these velocity power spectra to the perturbative

calculation as discussed in Section (3.1.1-3.1.3) is to perform a numerical study based on N-body simulations and find an empirical relation to model the velocity power spectra. [Bel et al. \(2019\)](#) performed study based on a set of N-body simulation Dark Energy and Massive Neutrinos Universe (DEMNUi) ([Carbone et al., 2016](#)) in the presence of massive neutrino. The velocity field was reconstructed from the cold dark matter particles using a Delaunay tessellation. The paper provided a fitting formula for the velocity power spectra as

$$P_{\delta\theta}(k) = \left\{ P_{\delta\delta}(k) P_{\theta\theta}^{\text{Lin}}(k) \right\}^{\frac{1}{2}} e^{-\frac{k}{k_\delta} - b k^6}, \quad (3.25)$$

$$P_{\theta\theta}(k) = P_{\theta\theta}^{\text{Lin}}(k) e^{-k(a_1 + a_2 k + a_3 k^2)}, \quad (3.26)$$

in our case, the input nonlinear matter power spectrum can be either calculated from RE-SPRESSO or from HALOFIT. The relation between  $P_{\delta\theta}$ ,  $P_{\theta\theta}$  is far from universal, but very much influenced by the amplitude of the matter fluctuation. The free parameters that enter Eq. (3.25) and Eq. (3.26) are given by

$$\begin{aligned} a_1 &= -0.817 + 3.198\sigma_{8,m} \\ a_2 &= 0.877 - 4.191\sigma_{8,m} \\ a_3 &= -1.199 + 4.629\sigma_{8,m} \\ 1/k_\delta &= -0.017 + 1.496\sigma_{8,m}^2 \\ b &= 0.091 + 0.702\sigma_{8,m}^2 \\ 1/k_\theta &= -0.048 + 1.917\sigma_{8,m}^2, \end{aligned} \quad (3.27)$$

here  $\sigma_{8,m}$  is the total matter fluctuation, including the cold dark matter as well as the massive neutrino. [Bel et al. \(2019\)](#) showed that these fitting function can provide an accuracy of  $\sim 3\%$  accuracy for  $P_{\delta\theta}$  for  $k < 0.7 h\text{Mpc}^{-1}$  and  $P_{\theta\theta}$  for  $k < 0.65 h\text{Mpc}^{-1}$  at redshift down to  $z = 0$ .

Fig. 3.2a shows the comparison for the matter-velocity cross power spectrum  $P_{\delta\theta}$  obtained from fitting function (see Eq. (3.25)) and the perturbative calculation from gRPT (blue) at two redshift snapshots  $z = \{0.57, 1.0\}$ . To obtain  $P_{\delta\theta}$  we input the matter power computed from RESPRESSO (red). Fig. 3.2b shows the ratio between the two approaches with the grey band being 3% error bar. Fig. 3.2c shows the auto power spectrum of the velocity divergence field  $P_{\theta\theta}$  given by the fitting formula (see Eq. (3.26)) and the gRPT calculation. Under the linear approximation,  $P_{\delta\delta}^{\text{lin}} = P_{\delta\theta}^{\text{lin}} = P_{\theta\theta}^{\text{lin}}$ . We input linear matter power spectrum  $P_{\delta\delta}^{\text{lin}}$  to calculate  $P_{\theta\theta}$  and  $P_{\delta\delta}^{\text{lin}}$  is calculated from CAMB ([Lewis et al., 2000](#)). Fig. 3.2d is the ratio between the fitting function approach and the gRPT calculation, with the grey band



being the 3% error bar. A good agreement is observed between the empirical fitting formula and the perturbative calculation by gRPT for  $P_{\delta\theta}$  and  $P_{\theta\theta}$ . Due to the nonlinear correction, the velocity power is suppressed in comparison to the amplitude of the matter power in Fig. 3.1.

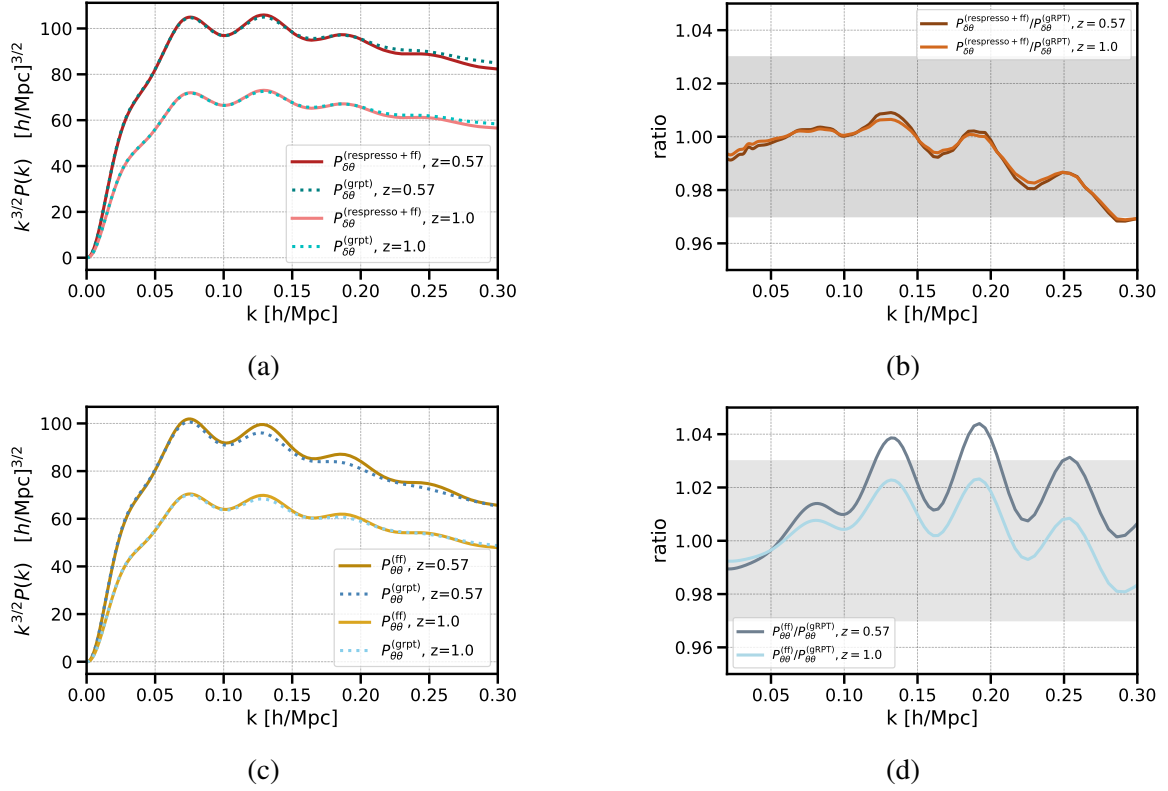


Fig. 3.2 Comparison of cross power spectrum for the matter-velocity divergence  $P_{\delta\theta}$  and the auto power spectrum for the velocity divergence  $P_{\theta\theta}$ . Fig. 3.2a shows  $P_{\delta\theta}$  at two redshift snapshots  $z = \{0.57, 1.0\}$  for the fitting function (see Eq. 3.25) with the input matter power spectrum computed from RESPRESSO (red) and the perturbative calculation from gRPT. Fig. 3.2b shows the ratio between the two approaches with the grey band showing 3% error bar. Fig. 3.2c shows  $P_{\theta\theta}$  given by the fitting formula (see Eq. 3.26) and the gRPT calculation, where the linear matter power spectrum is calculated from CAMB (Lewis et al., 2000). Fig. 3.2d is the ratio between the fitting function approach and the gRPT calculation, with the grey band being the 3% error bar.

## 3.2 Large scale structure bias

The large scale structure bias represents the statistical relation between the distribution of the luminous tracers and the underlying matter field. Down to the quasi-linear scales, the

statistical relation can be described as a perturbative bias expansion, which comprehend complicated galaxy formation process dominated by the local gravitational effects. The perturbative expansion of the galaxy density fluctuation  $\delta_g$  in terms of the matter fluctuation  $\delta$  can be generalized as (Desjacques et al., 2018),

$$\delta_g(\mathbf{x}, \tau) = \sum_O b_O(\tau) O(\mathbf{x}, \tau), \quad (3.28)$$

with  $O$  being the operators. The operators can be the matter fluctuation at a given point in space, or other quantities that affect the galaxy formation process.

### 3.2.1 Local and nonlocal bias

One interesting question arises as we expand the galaxy over-density in terms of the operator, what is the most efficiently described in a basis where the operators are linearly independent at each order? We assume the galaxies and the dark matter are comoving, i.e. there is no velocity bias, Galilean invariance of the equation of motion implies only the second derivative of the velocity potential is allowed. Consider all scalar invariants of the tensor  $\nabla_{ij}\Phi(\mathbf{x}, \tau)$  and  $\nabla_{ij}\Phi_v(\mathbf{x}, \tau)$ , the Cayley-Hamilton theorem (Hoffman & Kunze, 1971) states that only three such invariants exists in three dimension. These invariants are the so called Galileons operators,

$$\mathcal{G}_1(\Phi) \equiv \nabla^2 \Phi \equiv \delta, \quad (3.29)$$

$$\mathcal{G}_2(\Phi) \equiv \left( \nabla_{ij} \Phi \right)^2 - \left( \nabla^2 \Phi \right)^2, \quad (3.30)$$

$$\mathcal{G}_3(\Phi) \equiv \left( \nabla^2 \Phi \right)^3 + 2 \nabla_{ij} \Phi \nabla_{jk} \Phi \nabla_{ki} \Phi - 3 \left( \nabla_{ij} \Phi \right)^2 \nabla^2 \Phi, \quad (3.31)$$

similar relations also exist for the velocity potential  $\nabla^2 \Phi_v(\mathbf{x}, \tau) \equiv \theta(\mathbf{x}, \tau)$  and Eq. (3.2.1) can be associated with the tidal field. At linear order, the gravitational potential,  $\Phi$ , and velocity potential,  $\Phi_v$ , are equal to each other and the bias relation is dominated by the local processes directly given by the Galileons. At non-linear order the two potential terms are not equal to each other, as is shown in Chan et al. (2012), an additional operator emerges from the second Galileon operator  $\mathcal{G}_2^{(3)}$  at the third order,

$$\Delta_3 \mathcal{G}_2 = \mathcal{G}_2(\Phi) - \mathcal{G}_2(\Phi_v) = \mathcal{G}_2^{(3)}(\Phi) - \mathcal{G}_2^{(3)}(\Phi_v), \quad (3.32)$$

where the second and third order of  $\mathcal{G}_2$  is expanded with  $\Phi_v = \Phi_v^{(1)} + \Phi_v^{(2)}$ ,

$$\mathcal{G}_2^{(2)}(\Phi_v) = \left( \nabla_{ij} \Phi_v^{(1)} \right)^2 - \left( \nabla^2 \Phi_v^{(1)} \right)^2 \quad (3.33)$$

$$\mathcal{G}_2^{(3)}(\Phi_v) = 2 \left( \nabla_{ij} \Phi_v^{(1)} \nabla_{ij} \Phi_v^{(2)} - \nabla^2 \Phi_v^{(1)} \nabla^2 \Phi_v^{(2)} \right). \quad (3.34)$$

Further we use the relation (Eggemeier et al., 2019),

$$\mathcal{G}_1^{(2)}(\Phi) - \mathcal{G}_1^{(2)}(\Phi_v) = \delta^{(2)}(\mathbf{x}) - \theta^{(2)}(\mathbf{x}) = -\frac{2}{7} \mathcal{G}_2(\Phi_L), \quad (3.35)$$

$$\Phi^{(2)} - \Phi_v^{(2)} = -\frac{2}{7} \nabla^2 \mathcal{G}_2(\Phi_L), \quad (3.36)$$

where  $\Phi_L$  is the linear gravitational potential. The right hand side of Eq. (3.32) can be further rewritten using Eq. (3.34), (3.35) and (3.36) as

$$\mathcal{G}_2^{(3)}(\Phi) - \mathcal{G}_2^{(3)}(\Phi_v) \equiv \frac{4}{7} \left( \nabla_{ij} \Phi_v^{(1)} \nabla_{ij} \Phi_{2\text{LPT}} - \nabla^2 \Phi_v^{(1)} \nabla^2 \Phi_{2\text{LPT}} \right), \quad (3.37)$$

where  $\Phi_{2\text{LPT}}$  is the 2LPT potential for the displacement field to the second order with the Poisson equation  $\nabla^2 \Phi_{2\text{LPT}} = -\mathcal{G}_2(\Phi_v^{(1)})$ . Follow the bias expansion relation given in Chan et al. (2012), we expand galaxy density fluctuation field  $\delta_g$  as

$$\delta_g = b_1 \delta + \frac{b_2}{2} \delta^2 + \gamma_2 \mathcal{G}_2 + \gamma_3^- \Delta_3 \mathcal{G}_2 + \dots \quad (3.38)$$

To summarize, the operators can be split into two groups, the ones follow the local evolution, which can be directly built out the Galileons and the ones follow nonlocal evolution. Here we include the bias expansion up to the third order,

Local:	Non-Local:	
<b>1st:</b> $b_1 \delta$	<b>3rd:</b> $\gamma_3^- \Delta_3 \mathcal{G}$	(3.39)
<b>2nd :</b> $b_2 \delta^2, \gamma_2 \mathcal{G}_2$		

### 3.2.2 Higher-derivative bias

One type of large scale structure bias we do not include in our analysis is the so called higher-derivative bias, which involve operators higher than the second derivative of the gravitational and velocity potential. Such bias can in principle become important at scales, where the galaxy formation process becomes important (Desjacques & Seljak, 2010). For galaxies, such scales is usually given by the as Lagrangian radius  $R(M)$ , with the Lagrangian radius of the host halo with mass  $M$ ,  $R(M) = (3M/4\pi\bar{\rho}_m)^{1/3}$ . Below such scales, the galaxy formation due to the matter collapsing can no longer be approximated as local processes and the bias expansion is needed to include operators at multi spatial points. A typical term is  $\nabla^2 \delta$  and correspondingly  $k^2 \delta$  in Fourier space. Beyond the Lagrangian radius, such term is suppressed by the power  $(kR)^2$ . Since the minimum fitting scales is  $s_{\min} = 20\text{Mpc } h^{-1}$ ,

we think *a priori* such contribution is not important. Indeed, the modelling can recover the fiducial value of the N-body simulation very well. However, since the AGN is complicated, it would be interesting to include more tests in this direction.

### 3.3 Nonlinear modelling of the RSD

#### 3.3.1 Nonlinear RSD on large scales

We show in Section 2.3.2 as a heuristic derivation that, the RSD is directly related to the growth of the cosmic structure. The velocity is proportional to the density fluctuation field and the strength of the anisotropy in the clustering is given by the growth rate  $f$ . In the linear theory, the power spectrum in the redshift space is linked to the one in the real space by simply squaring both sides of Eq. (2.50). To further extract the information more accurately from the anisotropic clustering, we need to take the nonlinearity into account as we model the RSD. Starting with the definition, the power spectrum in the redshift space is given by

$$P^{(S)}(k) = \int d^3x e^{ik \cdot x} \left\langle e^{-ik\mu f \Delta u_z} \{ \delta(\mathbf{r}) + f \nabla_z u_z(\mathbf{r}) \} \times \{ \delta(\mathbf{r}') + f \nabla_z u_z(\mathbf{r}') \} \right\rangle, \quad (3.40)$$

where  $u_z = -v_z/(aHf)$  and  $\{ \delta(\mathbf{r}) + f \nabla_z u_z(\mathbf{r}) \}$  is the mapping between the density fluctuation in the real space to the redshift space (see Eq. (2.50)). Rename the variables and write the above equation in a more compact form,

$$P^{(S)}(k, \mu) = \int d^3x e^{ik \cdot x} \left\langle e^{\lambda O_1} O_2 O_3 \right\rangle, \quad (3.41)$$

where the operators are defined as,

$$\begin{aligned} \lambda &= -ik\mu f, & O_1 &= u_z(\mathbf{r}) - u_z(\mathbf{r}') \\ O_2 &= \delta(\mathbf{r}) + f \nabla_z u_z(\mathbf{r}), & O_3 &= \delta(\mathbf{r}') + f \nabla_z u_z(\mathbf{r}'). \end{aligned} \quad (3.42)$$

Use the relation between the cumulant and moment generating function,

$$\left\langle e^{\lambda \cdot \mathcal{O}} \right\rangle = \exp \left\langle e^{\lambda \cdot \mathcal{O}} \right\rangle_c, \quad (3.43)$$

where  $\mathcal{O} = \{O_1, O_2, O_3\}$  and the same for  $\lambda$ . The angle bracket in Eq. (3.41) is evaluated by taking derivative of both sides of Eq. 3.43 with respect to  $\lambda_{2,3}$  and then set them to zero (Scoccimarro, 2004; Taruya et al., 2010),

$$\begin{aligned} \langle e^{\lambda O_1} O_2 O_3 \rangle &= \exp \langle e^{\lambda O_1} \rangle_c \times \left[ \langle e^{\lambda O_1} O_2 O_3 \rangle_c + \langle e^{\lambda O_1} O_2 \rangle_c \langle e^{\lambda O_1} O_3 \rangle_c \right] \\ &\approx \exp \langle e^{\lambda O_1} \rangle_c \left\{ \langle O_1 O_2 \rangle + \lambda \langle O_1 O_2 O_3 \rangle_c + \lambda^2 \langle O_1 O_2 \rangle_c \langle O_1 O_3 \rangle_c + \dots \right\} \end{aligned} \quad (3.44)$$

In the second equation the expressions are perturbatively expanded up to second order with respect to  $\lambda$  and drop the higher order contributions that involves  $\langle O_1^2 O_2 O_3 \rangle_c$ . The operators inside the curly brackets are directly related to the non-virial power spectrum  $P_{\text{novir}}$  in Eq. 3.1, which can be further split into three terms as

$$P_{\text{novir}}(\mu, k) = P_{\text{novir}}^{(1)}(k, \mu) + (k\mu f) P_{\text{novir}}^{(2)}(k, \mu) + (k\mu f)^2 P_{\text{novir}}^{(3)}(k, \mu) \quad (3.45)$$

where the first term  $P_{\text{novir}}^{(1)} = \langle O_1 O_2 \rangle$  and represents the nonlinear Kaiser formula,

$$P_{\text{novir}}^{(1)}(k, \mu) = P_{\text{gg}} + 2f\mu^2 P_{\text{g}\theta} + f^2\mu^4 P_{\theta\theta}, \quad (3.46)$$

the ingredients for calculating  $P_{\text{gg}}$ ,  $P_{\text{g}\theta}$ , and  $P_{\theta\theta}$  are discussed in Section 3.1 and Section 3.2. The second term involves the bispectrum  $P_{\text{novir}}^{(2)} = \langle O_1 O_2 O_3 \rangle_c$  and is given by

$$P_{\text{novir}}^{(2)}(k, \mu) = \int d^3p \frac{p_z}{p^2} [B_\sigma(\mathbf{p}, \mathbf{k} - \mathbf{p}, -\mathbf{k}) - B_\sigma(\mathbf{p}, \mathbf{k}, -\mathbf{k} - \mathbf{p})], \quad (3.47)$$

where the cross bispectra calculated from the tree-level PT for the density and velocity including bias parameters for  $b_1$ ,  $b_2$  and  $\gamma_2$ ,

$$\begin{aligned} &\left\langle \theta(\mathbf{k}_1) \left\{ \delta_g(\mathbf{k}_2) + f \frac{k_{2z}}{k_2^2} \theta(\mathbf{k}_2) \right\} \left\{ \delta_g(\mathbf{k}_3) + f \frac{k_{3z}}{k_3^2} \theta(\mathbf{k}_3) \right\} \right\rangle \\ &= (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\sigma(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \end{aligned} \quad (3.48)$$

The third term is of quadratic order in the power spectrum and corresponds to  $P_{\text{novir}}^{(3)} = \langle O_1 O_2 \rangle_c \langle O_1 O_3 \rangle_c$

$$P_{\text{novir}}^{(3)}(k, \mu) = \int d^3p \frac{p_z (k_z - p_z)}{p^2 (k - p)^2} (b_1 + f\mu_p^2) (b_1 + f\mu_{k-p}^2) P_{\delta\theta}(\mathbf{p}) P_{\delta\theta}(\mathbf{k} - \mathbf{p}). \quad (3.49)$$

### 3.3.2 Nonlinear RSD at small scale: Finger of God factor

On small scales the random motion of LSS smears the distribution along the line of sight direction and give rise to the so called finger of god (FoG) effect. The term can be recognized

from the exponential factor in Eq. 3.44. It is common to model the FoG effect using a Gaussian or a Lorentzian function. However, as can be seen in Fig. 3.3 that the probability density distribution (PDF) for pairwise velocity along the line of sight direction is non-Gaussian. The pairwise velocity PDF is measured from the OuterRim simulation (Heitmann et al., 2019) (see next chapter, Section 4.4). The blue histogram uses the particles as satellites and the orange histogram assumes the satellite distribution follows the Navarro–Frenk–White (NFW) profile and we chose one separation slice for  $ds = 15 - 16 \text{ Mpc } h^{-1}$ . In both cases, the central galaxies are placed at the the most bound particle of the halo and inherit its velocity. We can see a clear difference between the velocity PDF and a Gaussian fit (dashed blue), which motivates to adopt the following function for the FoG factor,

$$F_{\text{FOG}}(\mu, k) \equiv \frac{1}{\sqrt{1 + \mu^2 k^2 a_{\text{vir}}^2}} \exp\left(\frac{-\mu^2 k^2 \sigma_v^2}{1 + \mu^2 k^2 a_{\text{vir}}^2}\right), \quad (3.50)$$

with  $a_{\text{vir}}$  being a free parameter that represents the kurtosis of the small scale velocity distribution. The one-dimensional linear velocity dispersion is given by  $\sigma_v^2 = \frac{1}{3} \int \frac{P_{\theta\theta}^{\text{lin}}(k)}{k^2} d^3 k$ . Such a treatment of FoG factor, which takes into account the nonlinear corrections, can also be found in Sánchez et al. (2017a), Grieb et al. (2017), and Hou et al. (2018). We do not explicitly express the  $f$  dependence as we described this before. Since there is a degeneracy in  $f$  and  $\sigma_8$ , we fit the combination of these two parameters. In a more general case, expressing the growth rate measurements in terms of  $f\sigma_8$  might lead to a bias, as the results depend on the assumed value of  $h$ , and an underestimation of the uncertainties, as the true constraints on  $f\sigma_8$  should be marginalized over the uncertainties on  $h$ . These problems are solved if the growth rate measurements are expressed instead in terms of the combination  $f\sigma_{12}$ , where  $\sigma_{12}$  represents the root-mean-square (rms) of the mass contained in a sphere of radius 12 Mpc (further see Chapter 4).

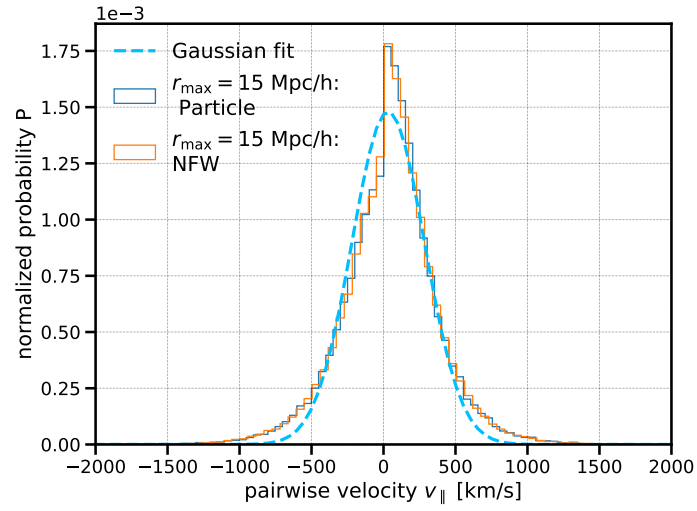


Fig. 3.3 the probability density distribution (PDF) for pairwise velocity along the line of sight direction separated by  $ds = 15 - 16 \text{ Mpc } h^{-1}$ . Blue histogram: satellites placed by random selection of the dark matter particle. Orange histogram: assumes the satellite distribution follows the Navarro–Frenk–White (NFW) profile. Blue dashed curve





## Chapter 4

# Clustering analysis of the eBOSS QSO

### 4.1 Data

#### 4.1.1 Overview of the eBOSS survey on QSOs

eBOSS is a five-year cosmological program as part of the SDSS project (see Section 1.3). eBOSS is the first survey that uses QSOs as tracers and successfully detected the BAO signals in the LSS ([Ata et al., 2017](#)). Quasars, whose luminosities powered by supermassive black holes at their centres, are intrinsically much more luminous than galaxies and can be detected at higher redshifts. Thus, they open a new redshift range for large scale structure clustering analyses. Our focus in the clustering analysis are the QSOs at redshift  $\sim 1.5$ . A sample of QSO for clustering analyses must be homogeneously selected. The design goal was to reach a survey area larger than  $7500 \text{ deg}^2$  with more than 435,000 objects, with less than 1% catastrophic redshift failure objects, and a maximum absolute variation in expected target density less than 15% for imaging survey sensitivity, stellar density and Galactic extinction ([Myers et al., 2015](#)). In the end, we were able to obtain 343,708 objects for clustering study. Table 4.1 summarizes a few properties of the quasar sample including the total number of objects and the weighted area for the north galactic cap (NGC) and south galactic cap (SGC). There are two steps to achieve the quasar data. In the first step, the quasar candidates are "targeted" from imaging data and in the second step these candidates go through a spectroscopic selection. In this section, we will go through these two steps and describe the related systematics.

Table 4.1 Summary of statistics for the eBOSS DR16 QSOs clustering catalogue. The number of eBOSS objects and Legacy objects (details see Section 4.1.3) are obtained with a redshift cut  $0.8 < z < 2.2$ , completeness  $C_{\text{eBOSS}} > 0.5$  and sector success rate  $C_z > 0.5$ .

	NGC	SGC	Total
$N_{\text{eboss}}$	173,838	109,437	283,275
$N_{\text{legacy}}$	44,371	16,062	60,433
$N_{\text{cp}}$	6878	4832	11,710
Effective volume ( $[\text{Gpc}/h]^3$ )	0.120	0.065	0.185
Area (weighted, $\text{deg}^2$ )	2860	1839	4699

## 4.1.2 Imaging phase

### QSO target selection

The eBOSS QSO target selection uses data from both SDSS I/II/III and the Wide Field Infrared Survey Explorer (Wise; [Wright et al., 2010](#)). Although eBOSS does not add new imaging data to the previous data release, it updated the calibration technique and uses the so called "uber-calibration" [Padmanabhan et al. \(2008\)](#) that reduced the residual systematic error on all photometric bands ([Finkbeiner et al., 2016](#)). The "Extreme Deconvolution" algorithm (XDQSO) was introduced to obtain a most uniform quasar sample. For eBOSS the algorithm is improved to XDQSOz ([Bovy et al., 2012](#)) that can be applied to any redshift range. The eBOSS QSO target selection also applies a mid infrared cut using the data obtained from WISE survey for band W1 ( $3.4\mu\text{m}$ ) and W2 ( $4.6\mu\text{m}$ ). The WISE data helps to distinguish quasars from stellar objects, since AGNs usually have extensive IR emission while the stars do not. The quasars that satisfy the target selection but do not have a previously known and secure redshift are denoted as "CORE" quasar **QSO\_CORE**. Combining the optical (by XDQSOz) and mid-IR cut, the final eBOSS CORE quasar target selection is

$$\begin{aligned} \text{PQSO}(z > 0.9) &> 0.2, \\ m_{\text{opt}} - m_{\text{IR}} &\geq (g - i) + 3, \end{aligned} \tag{4.1}$$

where the first criteria corresponds to the probability of being a quasar at  $z > 0.9$  higher than 20%. Here the point sources are selected with de-extincted PSF magnitudes with  $g < 22$  or  $r < 22$ . The second criteria corresponds to a selection of targets with significantly larger optical magnitude than IR magnitude, where the magnitude is converted from the stacked optical and WISE infrared flux given by equation (1) and equation (2) in [Myers et al. \(2015\)](#).

### Imaging systematics

One of the most important sources for the clustering systematics is the inhomogeneity in the photometry. Fig. 4.1 (Image credit: [Ross et al. \(2020\)](#)) shows the ratio of the quasar number density to the randoms  $n_{\text{QSO}}/n_{\text{random}}$  as a function of the imaging properties. The black dashed curve shows the raw data in which we can observe obvious trends with the  $g$ -band depth, Galactic extinction  $E(B - V)$ , sky background, and seeing in  $i$ -band.

During the eBOSS DR14 analysis, we identified  $g$ -band depth, Galactic extinction  $E(B - V)$  to be the important imaging systematics, where the  $g$ -band depth is the  $5\text{-}\sigma$  detection in magnitude (also called depth). For each point source in SDSS, it can be defined for each filter as,  $m_{5\sigma} = -2.5 \log \left( A \cdot \text{PSF}_{\text{FWHM}} \cdot \sqrt{\Phi_{\text{sky}}} 10^{0.4k \text{Airmass}} \right) - m_{\text{ext}}$ , with  $A$  being an overall scaling coefficient,  $\text{PSF}_{\text{FWHM}}$  is defined as the full width half maximum (FWHM) of the point-source function,  $\Phi_{\text{sky}}$  is the observed sky flux without the point sources,  $k$  is the atmospheric extinction coefficients, and  $m_{\text{ext}}$  is a correction to the Galactic extinction. The parameters  $A$  and  $k$  depends on the five bands and is different for NGC and SGC. During DR14, we corrected for the dependency in depth and the Galactic extinction, hence the total systematic weights composes two parts

$$w_{\text{sys}} = \frac{1}{(A_d + dB_d)(A_e + eB_e)}, \quad (4.2)$$

a linear fit is iteratively applied to the extinction corrected  $g$  band depth and the the Galactic Extinction in terms of  $n_{\text{QSO}}/n_{\text{random}}$  ratio. For the final data release, we also correct for the trends in the seeing and sky background following a same spirit. As can be seen in Fig. 4.1 that, all trends are reduced after applying the correction (green error bar).

### 4.1.3 Spectroscopic phase

#### Spectroscopic observation

After the quasars are targeted, the next step is to obtain the redshifts for each quasar. At the spectroscopic phase, first the tiling algorithm (a scheme to allocate the fibers) is designed to optimize the number of targets with minimum number of tiles. The fibers are plugged into the Aluminum plates (right panel in Fig. 1.3). Each plate subtends  $7 \text{ deg}^2$  ( $3^\circ$  diameter). The minimum angular projected distance between two neighbouring quasar targets in each observation is limited by the ferrules (a small bracelet) that supports the fibres, which have a physical size of  $62''$ . The tiling process aims at minimizing the number of objects falling within  $62''$  and maximizing the fraction of the fibers used for the high priority targets. The LSS quasars probe a completely new redshift range and are intrinsically shot noise dominated,

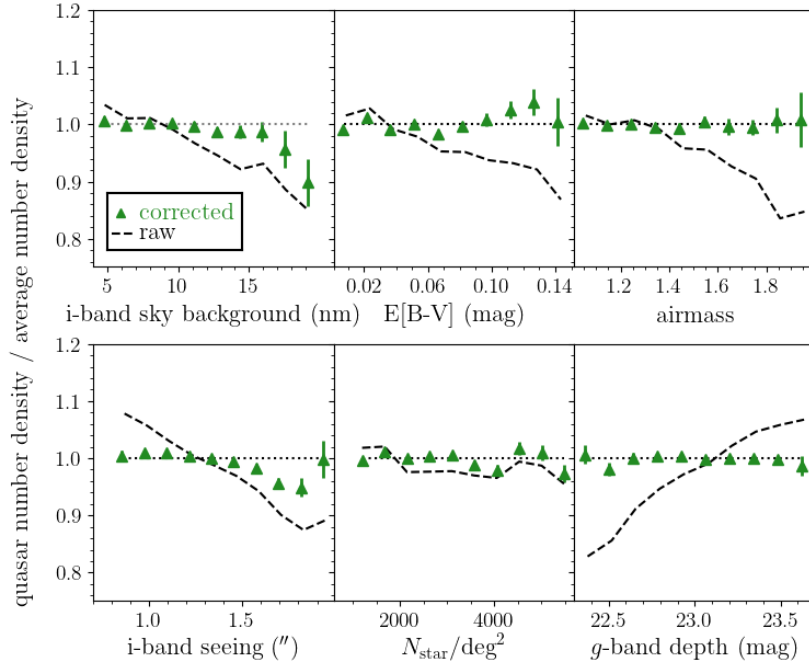


Fig. 4.1 Fluctuations in projected quasar density as a function of various image properties and Galactic foregrounds, combining NGC and SGC results (but normalizing them separately). The dashed curves show the result before weights for  $g$ -band depth and  $E(B - V)$  are applied. Image credit: [Ross et al. \(2020\)](#)

therefore, during the tiling process the "CORE" sample acquires almost always the highest priority compared to other targets. Fig. 4.2 shows the footprint for the final DR16 QSOs, where the left panel shows the north galactic cap (NGC) and the right plot shows the southern galactic cap (SGC). The color map denotes the completeness weighted by sector. In DR16 a completeness of  $\sim 98\%$  is achieved. The secured redshift sources can be divided into three classes, 1). **Legacy**: These quasars with reliable redshifts were obtained during SDSS I/II/III. Within this category, the objects observed before BOSS are obtained from a combination of a fifth edition of SDSS QSO catalog (based on SDSS DR7) ([Schneider et al., 2010](#)) and a catalog of known stellar spectra from SDSS-I/II. While eBOSS program does not allocate fiber to those objects with confident spectroscopic classification and a reliable redshift. 2). **SEQUELS**: Prior to the eBOSS, the Sloan Extended Quasar, ELG, and LRG Survey (SEQUELS) was designed as a pilot survey for eBOSS. The SEQUELS consisted in total 117 plates, 66 of which was observed during the end phase of the BOSS program (SDSS-III). During the first year of the eBOSS program added another 51 plates and these data were released in DR13 ([Albareti et al., 2017](#)). The SEQUELS used a less constrained quasar

selection algorithm than that adopted for eBOSS, therefore not all the targets meet the criteria by the eBOSS final selection and a subsample (still a large fraction) of the SEQUELS objects entered the final eBOSS catalogs. 3). **eBOSS**: During DR14, over 75 percent of the new redshifts were observed during the eBOSS program. In the final data release, this number has increased to  $\sim 82\%$ . A comparison of the footprint can be seen in Fig. 4.3. For the final data release, we increase both the survey area as well as the number of objects by a factor of 2. Fig. 4.4 shows the number density as a function of redshift for DR16 QSO for NGC (blue) and SGC (orange) for the final sample.

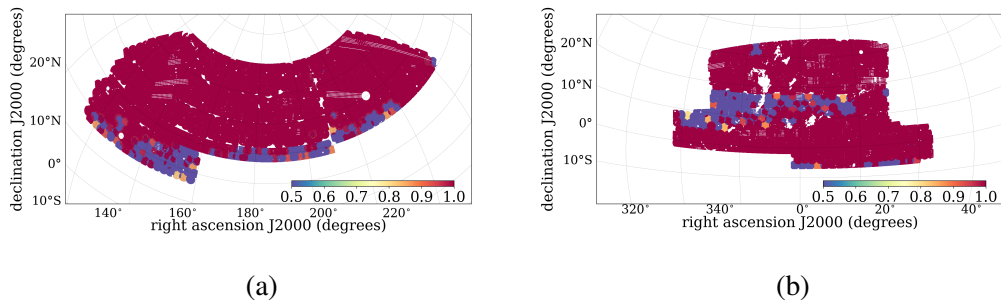


Fig. 4.2 Footprint of the DR16 QSO sample. The color coding is given by the completeness defined on each sector. The left panel shows the north hemisphere, the right panel shows the south hemisphere.

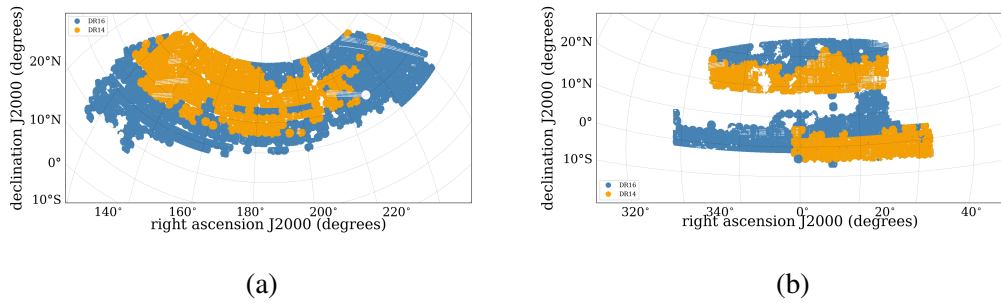


Fig. 4.3 Footprint of the eBOSS QSOs, split into the NGC (left) and SGC (right). The DR14 sample is shown in orange, while the DR16 sample is shown in blue (and also includes the entire orange region). The clustering catalogue only includes the objects with completeness  $\text{COMP\_BOSS} > 0.5$ .

### Spectroscopic systematics

During the spectroscopic observation, not all the objects can be successfully assigned with a redshift and one can accordingly calculate a redshift failure rate. The redshift failure rate

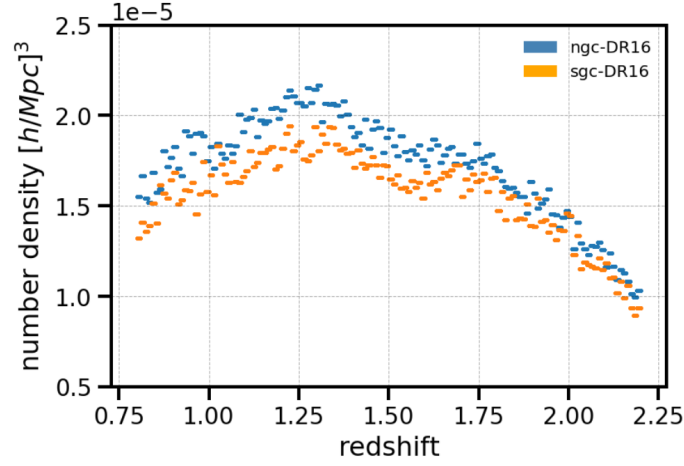


Fig. 4.4 The plot shows the number density as a function of the redshift of the eBOSS DR16 QSOs. The blue dots denote  $n(z)$  in NGC and the orange dots denote SGC.

can be correlated with the signal-to-noise ratio (SNR) of the quasar spectrum, as well as the observational condition, intrinsic brightness of the object, quality of the fiber. Fig. 4.5 shows the redshift efficiency as a function of fiber ID, where the black dots are the measured redshift efficiency on the data, while the red curve is a polynomial fit applied to the fiber efficiency for every 2 fibers with the edges at fiber ID = {1, 281, 501, 751, 1001}. The drop in

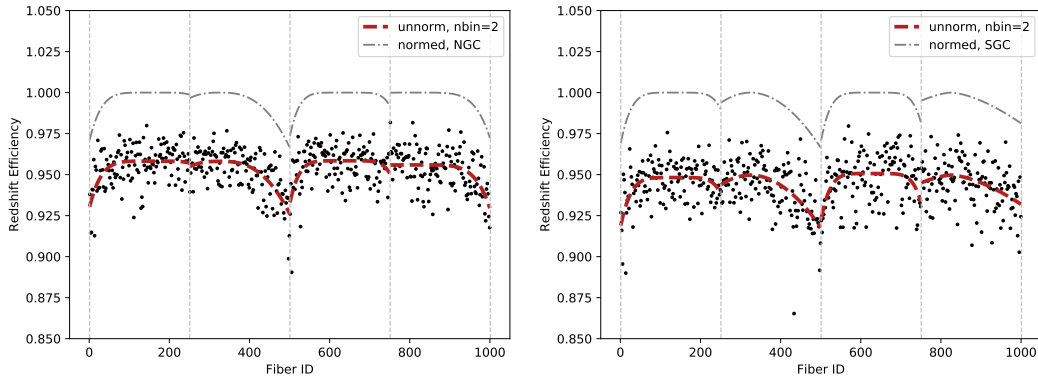


Fig. 4.5 Fiber efficiency as a function of fiber ID. A polynomial fit is applied to the fiber efficiency for every 2 fibers with the edges at fiber ID = {1, 281, 501, 751, 1001}. The left panel is for the north cap and the right panel is for the south cap.

redshift efficiency at fiber ID = {1, 501, 1001} corresponds to the quality of the pixels at the edge of the CCD cameras, while the drops at fiber ID = {281, 751} correspond to the location of the CCD amplifiers. To correct for the fiber ID dependent redshift efficiency rate, we cut the fiber ID into 4 regions and for each region a polynomial fit with to the redshift efficiency

is applied,

$$p = a - b|x - c|^d, \quad (4.3)$$

where  $\{a, b, c, d\}$  are 4 fitting parameters. The redshift efficiency is defined by,

$$r_{\text{zeff}}(\text{ID}) = \frac{N_{\text{obj}}[\text{IMATCH} = 1|4|9](\text{ID})}{N_{\text{obj}}[\text{IMATCH} = 1|4|7|9](\text{ID})}, \quad (4.4)$$

with  $[\text{IMATCH} = 1|4|9]$  being good eBOSS/SEQUELS redshifts, star and galaxies, respectively, while  $\text{IMATCH} = 7$  denotes the redshift failure objects. The weight is defined as the inverse of the redshift efficiency,  $w_{\text{noz},r} = r_{\text{zeff}}(\text{ID})^{-1}$ .

For quasars, the trend in the spectral signal to noise ratio (SNR) is very weak, which is likely due to the fact that the quasars are identified by the strong emission lines with higher SNR than the continuum. Nevertheless, we define weight  $w_{\text{noz},s}$  by performing a linear fit (more detail can be found in [Ross et al. \(2020\)](#)) to remove the residual spurious trends that could cause unphysical clustering signals. The final  $w_{\text{noz}}$  composes of the two parts mentioned above:  $w_{\text{noz}} = w_{\text{noz},s} w_{\text{noz},r}$ .

As mentioned before, the distance between two detectable neighbouring objects during one observation is limited by the ferrules (a small bracelet) that supports the fibers, with a projected size of  $62''$ . When two objects fall within such angular separation, they are denoted as "collided objects" and corrected using the close pair (fiber collision) weight  $w_{\text{cp}}$  by up-weighting the objects according to the colliding fraction of each group.

Finally, we reach at a series of weights to correct for the observing, targeting systematics on the quasar density contrast. The final weight applied to the objects is defined by

$$w_{\text{tot}} = w_{\text{FKP}} \cdot w_{\text{sys}} \cdot w_{\text{cp}} \cdot w_{\text{noz}}, \quad (4.5)$$

where the radial weight ([Feldman et al., 1994](#))  $w_{\text{FKP}} = (1 + P_0 n(z))^{-1}$  is applied to minimise the variance of measurement, with  $P_0 = 6000 h^{-3} \text{Mpc}^3$  and  $n(z)$  is the volume number density at each redshift bin.

The complicated quasar physics poses challenges for the quasar redshift estimation. The emission lines from quasars can be not only broadened but also subject to the quasar outflow that lead to lower redshift resolution and systematic bias ([Hewett & Wild, 2010](#)). The low redshift resolution can further smear the quasar distribution along the line of sight direction in redshift space (see the pronounced FoG feature in right panel of Fig. 4.6). Fortunately, this effect can be largely mitigated within the RSD modelling. Line confusions and residuals from sky subtractions can lead to redshift errors that are classified as "catastrophic". During the DR14 analysis, several the redshift estimates were tested. The redshift estimates include one

that is obtained from the standard SDSS pipeline in combination with the visual inspection ( $\sim 7\%$ ), a Mg II emission line based estimate, and one based on the principal component analysis with a prior on Mg II emission line. The first option was used during the DR14 analysis, and all the redshift estimates were found to be in agreement within  $1\sigma$  in terms of cosmological parameters (Zarrouk et al., 2018). However, the pipeline version was optimized for the BOSS Ly $\alpha$  quasars at  $z > 2.2$ , and the catastrophic redshift failure rate for the pipeline version is estimated to be 2.1% (Lyke et al., 2020). The standard "Z" column used to be a mixture of the pipeline redshift and visual inspection.

For DR16 analysis we use a new algorithm REDVSBLUE based on principle component analysis (PCA)<sup>1</sup>, which was developed to fit for emission-line based redshift estimates. The new algorithm provides redshift estimates in a more consistent way that one does not need to worry about small systematic offset due to visual inspection done by different people.

## 4.2 Two point correlation function

The correlation function  $\xi(\mathbf{s})$  as defined in Eq. (2.32) characterizes the probability in excess of random of observing pairs of galaxies as a function of their separation,  $\mathbf{s}$ . Assuming rotational symmetry along the line of sight direction, the correlation function is reduced to the two-dimensional function  $\xi(\mathbf{s}) \equiv \xi(\mu, s)$ , where  $\mu = \cos(\theta)$ , and  $\theta$  is the angle between the separation vector  $\mathbf{s}$  and the line of sight direction. Fig. 4.6 shows two dimensional correlation function  $\xi(s_\perp, s_\parallel)$  measured from the final DR16 QSO sample. The analysis of the full two-dimensional correlation function  $\xi(\mu, s)$  poses two problems: low signal-to-noise ratio and the large size of its covariance matrix.

Fortunately, the information of the full anisotropic correlation function can be condensed into a small set of one-dimensional projections by choosing different angular-dependent weighting schemes. One of the typical choices is decomposing the correlation function into the Legendre polynomial  $L_\ell(\mu)$ ,

$$\xi(\mu, s) = \sum_{\ell} \xi_{\ell}(s) L_{\ell}(\mu), \quad (4.6)$$

using orthogonality of the Legendre polynomials  $\int_0^1 L_{\ell}(\mu) L'_{\ell'}(\mu) d\mu = \delta_{\ell, \ell'}$ , one arrives at

$$\xi_{\ell}(s) \equiv \frac{2\ell+1}{2} \int_{-1}^1 \xi(\mu, s) L_{\ell}(\mu) d\mu, \quad (4.7)$$

---

<sup>1</sup><https://github.com/londumas/redvsblue>



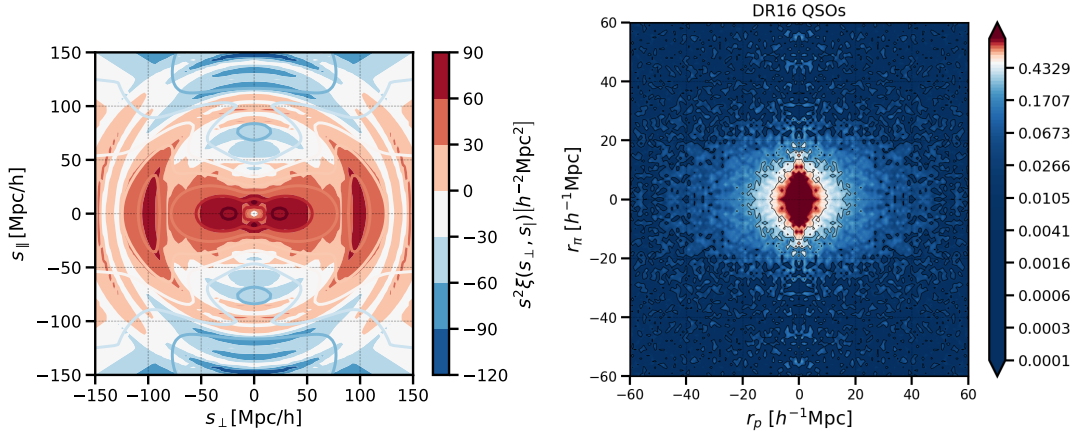


Fig. 4.6 Left: The 2D dimensional correlation function  $\xi(s_{\perp}, s_{\parallel})$  measured from DR16 quasar. The solid line is the theory prediction. Right: a zoom-in of the same plot at smaller scales.

due to the symmetry w.r.t  $\mu$ , only the even multipoles survive and the  $\ell = 0, 2, 4$  terms are referred to monopole, quadrupole and hexadecapole. Alternatively, one can replace the Legendre polynomial with 1 and bin over different angle ranges,

$$\xi_{\Delta\mu}(s) \equiv \frac{1}{\Delta\mu} \int_{\mu_1}^{\mu_2} \xi(\mu, s) d\mu, \quad (4.8)$$

this approach is commonly referred to as clustering wedges (Kazin et al., 2012), where  $\Delta\mu = \mu_2 - \mu_1$ . More specifically, the angle can be divided into two or three equal-width intervals  $\xi_{nw,i}(s)$ , with  $n = 2, 3$ , for the intervals  $(i-1)/n < \mu < i/n$ . The clustering wedges are broadly equivalent to the basis of Legendre polynomials, plug Eq. (4.6) into Eq. (4.8),

$$\xi_{\Delta\mu}(s) = \frac{1}{\Delta\mu} \int_{\mu_1}^{\mu_2} \sum_{\ell} \xi_{\ell}(s) L_{\ell}(\mu) d\mu = \xi_{\Delta\mu}(s) = \sum_{\ell} \xi_{\ell}(s) \bar{L}_{\ell}, \quad (4.9)$$

where  $\bar{L}_{\ell} = \frac{1}{\Delta\mu} \int_{\mu_1}^{\mu_2} L_{\ell}(\mu) d\mu$  is the average of the Legendre polynomial of order  $\ell$  over the  $\mu$ -bin of the clustering wedge. The transformation between three wedges to three multipoles,  $\bar{L}_{\ell=0,2,4}$ , can be explicitly written as

$$\bar{L}_{\ell} = \begin{bmatrix} 1 & -4/9 & 20/81 \\ 1 & -1/9 & -85/324 \\ 1 & 5/9 & 5/324 \end{bmatrix}, \quad \bar{L}_{\ell}^{-1} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -9/14 & -15/28 & 33/28 \\ 54/35 & -81/35 & 27/35 \end{bmatrix}. \quad (4.10)$$

Fig. 4.7 shows the Legendre multipoles  $\xi_{\ell=0,2,4}(s)$  (left panel) and the clustering wedges as a function of the pair separation with binning of  $ds = 8 h^{-1} \text{Mpc}$  measured from DR14

QSO sample. The upper panel in Fig. 4.7b displays the clustering wedges binned into three angle intervals, where the top wedge corresponds to the  $0 < \mu < 1/3$  (transverse direction), the middle wedge  $1/3 < \mu < 2/3$  and the down wedge  $2/3 < \mu < 1$  (parallel direction). The lower right panel is binned into two angle intervals. The error bars are inferred from the mock simulations described in Section 4.4.3. The BAO signal can be observed as a bump at scale  $ds \sim 110h^{-1}\text{Mpc}$  both for the monopole on the left panel and the  $\mu$ -wedges on the right panel. The dashed line in the figure corresponds to the best fit to the data points using the theoretical model described in Section 3.3 for RSD and Section 3.1.3 using the gRPT for matter power spectrum.

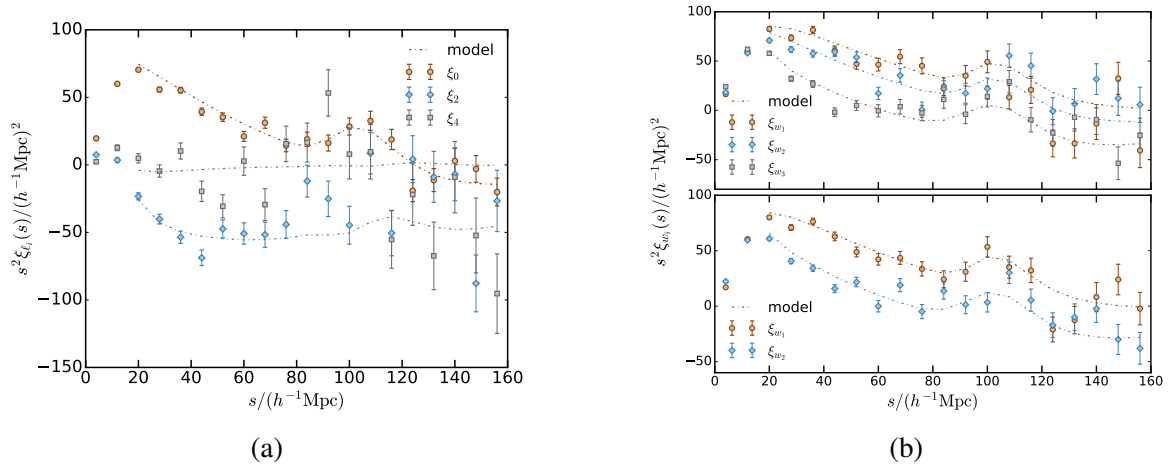


Fig. 4.7 Left: Legendre multipoles  $N_{\ell_i} = 3$ , monopole (red)  $\ell = 0$ , quadrupole (cyan)  $\ell = 2$ , and hexadecapole (grey)  $\ell = 4$ . Right: upper panel displays clustering wedges  $N_{w_i} = 3$  in the directions transverse (red) intermediate (cyan) and parallel (grey) to the line of sight and lower panel shows  $N_{w_i} = 2$  without the intermediate wedge. The multipoles and wedges are measured from the quasar sample of eBOSS DR14. The dashed lines correspond to the best fitting model to these measurements. The errorbars are inferred from  $10^3$  sets of mock catalogues (EZMOCKS).

Fig. 4.8 shows the measured two-point correlation function multipoles from the DR16 QSO sample with the best fitting models given by the solid curves. In order to highlight the BAO feature, in right panel of Fig. 4.8 the component of the best-fit model with no BAO has been subtracted. The bottom panel displays the result for the quadrupole. In order to highlight the (lack of) difference between  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ , we have subtracted the quadrupole of a model that has the same parameters as the best-fit, but with  $\epsilon = (\alpha_{\parallel}/\alpha_{\perp} - 1) = 0$ . If  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  differ,  $\epsilon \neq 0$ , a feature is observed in figure 3 from Alam et al. (2017). Here, we see that the BAO transverse and along the line of sight are consistent with each other with respect to our fiducial model.

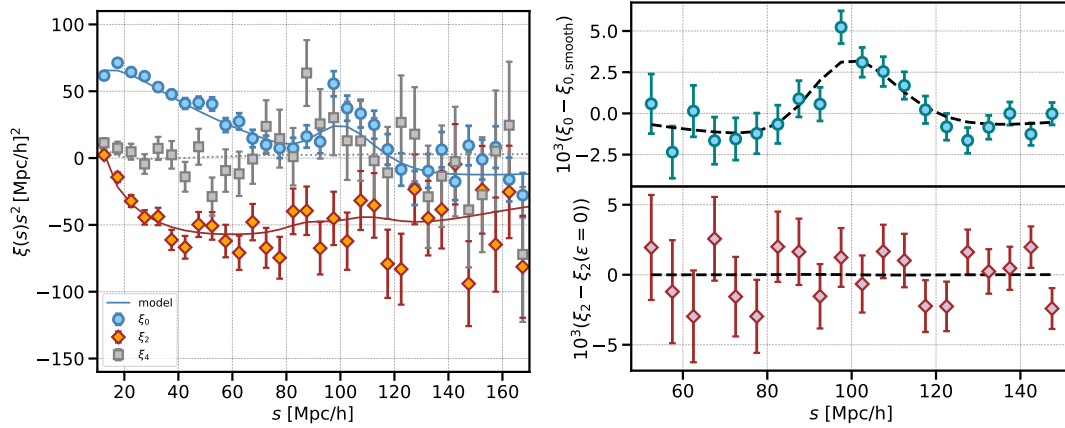


Fig. 4.8 Left: The measured correlation function for monopole ( $\ell = 0$ , blue), quadrupole ( $\ell = 2$ , red) and hexadecapole ( $\ell = 4$ , gray), with the best fitting full-shape model shown by the solid lines. Right: Comparison between our measured correlation function and the best-fit BAO model. In the top panel, we show the monopole, where we have subtracted the smooth component of the model from both the model and the data. In the bottom panel, we display the quadrupole and subtract the quadrupole of a model that has the same parameters as the best-fit, but with  $\epsilon = 0$ .

## 4.3 Methodology

### 4.3.1 Inference of cosmological parameters

In order to infer the best-fit cosmological parameters from a theoretical model, we aim to maximize the likelihood function. Given Bayes's theorem, the posterior distribution of a set of parameters  $\{\lambda\}$  is proportional to the product of the likelihood function and the prior  $\mathcal{P}(\lambda|\xi) \propto \mathcal{L}(\xi|\lambda)\mathcal{P}(\lambda)$ . In our case, the data vector  $\xi$  stands for the two-point correlation function. The likelihood for the Gaussian-distributed data is

$$\mathcal{L}(\xi|\lambda) \propto \exp \left[ -\frac{1}{2} (\xi - \xi_{\text{model}}(\lambda))^T \Psi (\xi - \xi_{\text{model}}(\lambda)) \right], \quad (4.11)$$

where  $\xi$  represents the data vector (more detail see Section 4.2), the precision matrix is the inverse of the true covariance matrix  $\Psi = \mathbf{C}_{\text{true}}^{-1}$ , which follows the inverse Wishart distribution (more discussion on the estimation of the covariance matrix see Section 4.3.2).  $\xi_{\text{model}}(\lambda)$  represents the theoretical model used to describe the two-point statistics.

In order to transform the observed redshift into distance, a fiducial cosmology is required. A difference between the true and fiducial cosmological parameters results in a rescaling of

cosmological distances. The geometric distortion parameters  $q_{\perp}$  and  $q_{\parallel}$  can be defined as

$$q_{\perp} = \frac{s_{\perp}}{s'_{\perp}} = \frac{D_M(z_m)}{D'_M(z_m)}, \quad (4.12)$$

$$q_{\parallel} = \frac{s_{\parallel}}{s'_{\parallel}} = \frac{D_H(z_m)}{D'_H(z_m)}, \quad (4.13)$$

with the prime ' denoting the distance inferred from the fiducial cosmology along and perpendicular to the line of sight  $s_{\parallel}$  and  $s_{\perp}$ . The rescaling of the 2D correlation function  $\xi(s, \mu) \rightarrow \xi(s', \mu')$  can be expressed as

$$s = s' \sqrt{q_{\parallel}^2 (\mu')^2 + q_{\perp}^2 (1 - \mu'^2)}, \quad (4.14)$$

$$\mu = \frac{q_{\parallel} \mu'}{\sqrt{q_{\parallel}^2 (\mu')^2 + q_{\perp}^2 (1 - \mu'^2)}}. \quad (4.15)$$

The BAO scale is tightly related to the comoving sound horizon at drag epoch,  $r_{\text{drag}}$ , that depends on the ratio of the baryon to radiation density. The geometric distortion parameters need further to be rescaled by the ratio of the sound horizon,

$$\alpha_{\perp} = q_{\perp} \frac{r'_{\text{drag}}}{r_{\text{drag}}} \quad \text{and} \quad \alpha_{\parallel} = q_{\parallel} \frac{r'_{\text{drag}}}{r_{\text{drag}}}, \quad (4.16)$$

where  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  are commonly referred to as the Alcock-Paczynski (AP) parameters ([Alcock & Paczynski, 1979](#)). This method of compressing the cosmological information is only an approximation, which we test in Section 5.1.2.

### 4.3.2 Covariance matrix estimation

The inference of the cosmological parameters by maximizing the likelihood requires an accurate covariance matrix. The covariance matrix can be estimated from a set of mock simulation (Section 4.3.2) or following an analytical approach (Section 4.3.2).

#### Covariance matrix from the mock catalogues

It is common to estimate the covariance from a large set of synthetic catalogues reproducing the properties of the true sample being considered. In this way, one can easily incorporate

the survey geometry into the covariance estimation. The covariance matrix is defined as

$$\hat{C}_*^{ij} = \frac{1}{n} \sum_{k=1}^{N_m} \left( \xi_i^{(k)} - \bar{\xi}_i \right) \left( \xi_j^{(k)} - \bar{\xi}_j \right), \quad (4.17)$$

where  $\xi_i^{(k)}$  is value for the  $k$ th mock at the  $i$ th data bin,  $\bar{\xi}_i$  is the expected value at the  $i$ th data bin. If the mean data vector  $\bar{\xi}$  is known,  $n$  equals to the number of synthetic mocks  $N_m$ , otherwise,  $n = N_m - 1$ .

However, the estimation of the covariance from a limited number of mocks poses two problems: firstly, the noise in the covariance will in general lead to a biased estimation of the precision matrix due to the inversion operation. Secondly, the noise in the covariance will propagate into the parameter constraints. Following [Anderson \(2003\)](#); [Hartlap et al. \(2007\)](#), we can correct for the bias in the precision matrix as

$$C_{\text{debiased}}^{-1} = \frac{n - N_b - 1}{n} \langle \hat{C}_*^{-1} \rangle \quad \text{for } N_b < n - 1, \quad (4.18)$$

where  $N_b$  represents the number of bins in the data vector. Regarding the impact of the covariance matrix noise on the parameter constraints, in [Dodelson & Schneider \(2013\)](#) it was shown that if the precision matrix is contaminated by the error  $\Psi = \Psi_{\text{true}} + \Delta\Psi$ , it leads to an additional term in the covariance when expanding the covariance to second order. When the best fitting parameters are estimated from a set of independent mock catalogues, the actual scattering of the best fitting parameters is inflated at the second order given by

$$\langle p_\alpha p_\beta \rangle \Big|_{\text{so.}} = B F_{\alpha\beta}^{-1} (N_b - N_p), \quad (4.19)$$

with

$$B = \frac{N_m - N_b - 2}{(N_m - N_b - 1)(N_m - N_b - 4)} \quad (4.20)$$

and the Fisher matrix  $F_{\alpha\beta}$ ,

$$F_{\alpha\beta} \simeq \sum_{ij} \frac{\partial x_i}{\partial p_\alpha} \Psi_{ij}^t \frac{\partial x_j}{\partial p_\beta}, \quad (4.21)$$

Fig 4.9 is a demonstration of the scattering of the best fitting parameter as a function of the number of synthetic simulation used to estimate the covariance matrix, where the covariance matrix is estimated from 10,000 PINOCCHIO mocks ([Monaco et al., 2013](#)). The total mocks are divided into two sets, the first 5,000 set is used to estimate the covariance and the second 5000 set is used to calculate the mean and used as the data to be fit. The blue curve is the prediction of the scattering of the best fitting parameter for a noisy covariance matrix

proposed in [Dodelson & Schneider \(2013\)](#). Here we use a toy model with a free parameter being the amplitude of the mean of the two-point correlation function. The inset of the figure shows the comparison between the mean amplitude and the true amplitude, the negligible difference implies that the covariance does not bias the parameter inference. At the same time we notice the standard deviation of the best fitting parameter decreases as the number of synthetic mocks increases, which agree very well with the theory prediction in the blue curve.

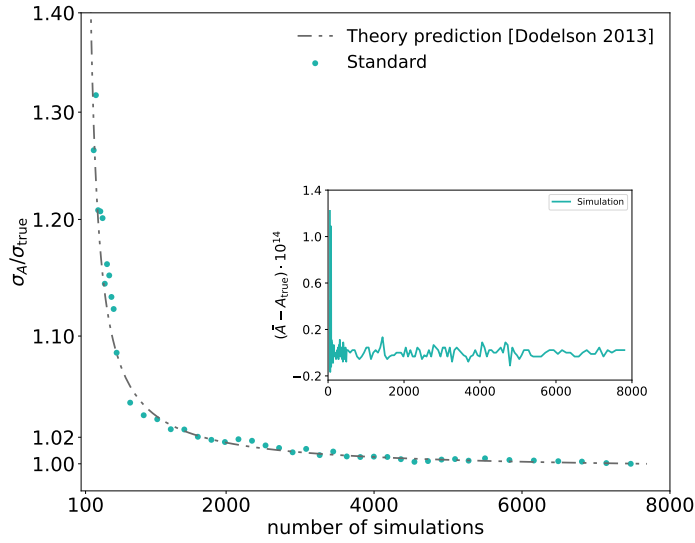


Fig. 4.9 Standard deviation of the best fitting parameter as a function of the number of synthetic simulation used to estimate the covariance matrix. The blue curve is the theory prediction proposed in ([Dodelson & Schneider, 2013](#)). The inset of the figure is a comparison between the free parameter in the toy model and the true value, which shows that there is essentially no bias in the inferred parameter.

Another factor we need to take into account is that, when infer the parameters from the real data, the error is actually derived by integrating the likelihood. Following [Percival et al. \(2014\)](#) the error in the covariance leads to a modified variance estimator

$$\hat{\sigma}_{\alpha\beta}^2 = [F + \Delta F]_{\alpha\beta}^{-1}. \quad (4.22)$$

Multiply both sides of Eq. 4.22 by  $F^{-1}$  and use the matrix property  $A^{-1}B^{-1} = [BA]^{-1}$  to expand  $\Delta F F^{-1}$ , we arrive at

$$\begin{aligned} F^{-1}\hat{\sigma}_{\alpha\beta}^2 &= F^{-1}[F + \Delta F]_{\alpha\beta}^{-1} \\ &= (1 + \Delta F F^{-1})_{\alpha\beta}^{-1} \\ &= (1 + \Delta F F^{-1} \Delta F F^{-1})_{\alpha\beta}. \end{aligned} \quad (4.23)$$

We assume that the covariance matrix is uncorrelated with its error, the expectation value of the first order vanishes  $\langle \Delta F F^{-1} \rangle = 0$ . Eq. (4.22) can be further rewritten as

$$\hat{\sigma}_{\alpha\beta}^2 = F_{\alpha\beta}^{-1} + \left( F^{-1} \Delta F F^{-1} \Delta F F^{-1} \right)_{\alpha\beta} \quad (4.24)$$

$$= \left[ 1 + A + B (N_p + 1) \right] F_{\alpha\beta}^{-1}, \quad (4.25)$$

where the parameter  $A$  is evaluated in [Percival et al. \(2014\)](#) and is given by

$$A = \frac{2}{(N_m - N_b - 1)(N_m - N_b - 4)}. \quad (4.26)$$

Eq. (4.25) indicates that the error derived from the likelihood are on average larger if the covariance matrix is contaminated by the noise, but this effect is only partially cancelled with the effect derived by [Dodelson & Schneider \(2013\)](#). As a conclusion, to obtain the parameter constraints inferred from the likelihood using a limited number of mock simulations, the final parameter covariance matrix needs to be rescaled by

$$M = \frac{1 + B (N_b - N_p)}{1 + A + B (N_p + 1)}. \quad (4.27)$$

### Analytical covariance matrix estimation

One problem associated with the brute force method is that for a large survey with high number density, it can be computationally very expensive to construct the required simulations. Another problem is that if the number of simulation does not fulfill  $N_s \gg N_b$ , the noise in the covariance matrix will propagate into the parameter estimation and the error bar can be overestimated. Furthermore, in some cases there are not enough mock catalogues to estimate the covariance matrix. Therefore, alternative methods such as analytical expression of the covariance can be very helpful. Following the prescription in ([Grieb et al., 2016](#)), the covariance for the Legendre multipoles in configuration space can be expressed as

$$C_{\ell_1 \ell_2}^{\xi}(s_i, s_j) = \frac{i_1^{\ell_1 + \ell_2}}{2\pi^2} \int_0^{\infty} k^2 \sigma_{\ell_1 \ell_2}^2(k) \bar{j}_{\ell_1}(ks_i) \bar{j}_{\ell_2}(ks_j) dk, \quad (4.28)$$

where the  $\bar{j}_{\ell}$  is the bin-averaged spherical Bessel function over a volume  $V_{s_i} = 4\pi (s_{i,\max}^3 - s_{i,\min}^3) / 3$  around a bin  $s_i$ ,

$$\bar{j}_{\ell}(ks_i) \equiv \frac{4\pi}{V_{s_i}} \int_{s_i - \Delta s/2}^{s_i + \Delta s/2} s^2 j_{\ell}(ks) ds. \quad (4.29)$$

The per-mode covariance in Eq. (4.28) is given by

$$\sigma_{\ell_1 \ell_2}^2(k) \equiv \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{V_{\text{eff}}} \times \int_{-1}^1 \left[ P(k, \mu) + \frac{1}{n} \right]^2 \mathcal{L}_{\ell_1}(\mu) \mathcal{L}_{\ell_2}(\mu) d\mu. \quad (4.30)$$

In the case of a simulation box the volume is estimated by  $V_{\text{eff}} = L^3$ , while for a survey with selection function and varying radial number density, the volume  $V_{\text{eff}}$  can be defined as the integral of the amplitude square of the survey window function  $Q(\mathbf{k})$  in Fourier space,

$$V_{\text{eff}}^{-1} \equiv \int \frac{d^3 \mathbf{k}}{(2\pi)^3} |Q(\mathbf{k})|^2 = \frac{\int d^3 \mathbf{x} \bar{n}^4(\mathbf{x}) w^4(\mathbf{x})}{\left[ \int d^3 \mathbf{x} \bar{n}^2(\mathbf{x}) w^2(\mathbf{x}) \right]^2}, \quad (4.31)$$

the window function is designed to maximize the signal-to-noise ratio by weighting the density fluctuation field. For mocks without survey geometry, the weight  $w(\mathbf{x})$  is simply the FKP weight

$$w(\mathbf{x}) = \frac{1}{1 + n(\mathbf{x}) P_{\text{FKP}}}, \quad (4.32)$$

with  $P_{\text{FKP}} = 6000$ . Here we have neglected the effects by the super-survey mode, such as the beat-coupling and the local average effect. The first effect is induced by the survey window that mixes the small- and large- scale modes. The second effect is caused by the zero-mode-modulated average density that is estimated from a limited survey. These two effects cancel each other and leave only upto  $\sim 10\%$  excess in the original variance (Putter et al., 2012). Within the fitting range we consider throughout our work, the analytical Gaussian covariance should remain a quite good approximation, which was also observed in (Lippich et al., 2019) when comparing to a set of fast mocks.

## 4.4 Assessing the systematic uncertainties

In this section, we will describe how we assess the systematic uncertainties. We split the systematic uncertainties into modelling and observational systematics.

To assess the modelling systematics, we perform a N-body mock challenge (Smith et al., 2020), using the OUTERRIM simulation (Heitmann et al., 2019). The OUTERRIM simulation was run in a cubic box of side length  $L_b = 3 h^{-1} \text{Gpc}$ , with  $10,240^3$  dark matter particles and a force resolution of  $6 h^{-1} \text{kpc}$ , corresponding to a mass resolution of  $m_p = 1.82 \cdot 10^9 \times M_\odot$ . The cosmology for the OUTERRIM simulation is consistent with WMAP7 cosmology (Komatsu et al., 2011), with  $h_0 = 0.71$ ,  $\Omega_b h^2 = 0.02258$ ,  $\Omega_{\text{cdm}} h^2 = 0.1109$ ,  $\sigma_8 = 0.8$ ,  $n_s = 0.963$ , and zero neutrino mass. The mocks are constructed from a cubic box using a single snapshot at



$z = 1.433$ , and are populated with quasars using halo occupation distribution (HOD) models. The goal of the mock challenge is two-fold: firstly, it serves to provide an estimate of the systematics in the modelling of the two-point statistics. Secondly, it is used to assess the impact of the assumption of the fiducial cosmology. In the first stage of the mock challenge (Section 4.4.1), we test our model on a ‘non-blind’ set of mocks, where we know precisely the underlying cosmology. In order to test the full analysis pipeline, in the second stage, we test our methodology on a set of ‘blind’ mocks which have been rescaled to different cosmologies. The true cosmological parameters of these mocks are unknown during the analysis (Section 4.4.2). The mock challenge is described in detail in [Smith et al. \(2020\)](#).

We quantify the observational systematics using a set of approximate EZMOCKS (Section 4.4.3). In the following sections, we summarize the tests we performed, and our main conclusions.

#### 4.4.1 Modelling systematics: non-blind mock challenge

The mock catalogues for the non-blind part of the mock challenge were created using 20 different halo occupation distribution (HOD) models, and we generate 100 random realizations of each. The HOD specifies the probability distribution for the number of galaxies inside a halo, where the halo mass is usually a dominating factor. The probability distribution is typically treated separately for the central galaxies and the satellite galaxies and the mean occupation number of a halo for a given mass can be expressed as

$$\langle N|M \rangle = \langle N_{\text{cen}}|M \rangle + \langle N_{\text{sat}}|M \rangle. \quad (4.33)$$

There are different ways to parametrize the mean occupation number for either the centrals or the satellites. The OUTERIM mocks built for QSOs (OR-QSOs mocks) utilize the options such as smooth step function for central galaxies with a power law for satellite galaxies (e.g. see [Tinker et al., 2012](#)), Gaussian for both central and satellite galaxies (e.g. see [Eftekharzadeh et al., 2019](#)). Apart from these options, the central galaxies are also placed using sharp step function and top hat function. In total 20 different parametrizations were assigned. The mock catalogues have been constructed so that they all have approximately the same clustering amplitude and number density comparable to the real data. For each of the 20 parametrizations, 100 realizations are generated from a single snapshot. On top of that, since the QSOs suffer from the redshift uncertainties as discussed in Section 4.1.3, therefore it is motivating to include such effects into the mocks. The mocks include 3 schemes of redshift distribution: a non-distorted one, a Gaussian smearing, and a realistic smearing. The uncertainty in a single redshift estimator is captured by the difference between the different

redshift estimators. One can mimic such effect by randomly drawing a redshift from a Gaussian distribution with a redshift dependent width as

$$\sigma_v(z) = \begin{cases} 300 \text{ kms}^{-1}, & \text{for } z \leq 1.5 \\ 400 \times (z - 1.5) + 300 \text{ kms}^{-1}, & \text{for } z > 1.5 \end{cases}. \quad (4.34)$$

Since the distribution of the redshift error exhibits a long tail (in fact, the distribution of the difference between redshift estimates, see figure 4 in [Zarrouk et al., 2018](#)), the shape can be described better with a double Gaussian distribution and we denote such distribution as the realistic smearing. For each realization, a random  $\Delta v$  is drawn from the following distribution

$$P(\Delta v) = N_1 \exp\left(\frac{-\Delta v^2}{2\sigma_1^2}\right) + N_2 \exp\left(\frac{-\Delta v^2}{2\sigma_2^2}\right), \quad (4.35)$$

where  $N_1 = 2000$ ,  $\sigma_1 = 150 \text{ kms}^{-1}$ ,  $N_2 = 67$ ,  $\sigma_2 = 1000 \text{ kms}^{-1}$ , the redshift uncertainty can be accordingly converted using the relation  $\Delta z = \Delta v(1+z)/c$  and added to each object.

Finally, in the real data it is difficult to completely avoid line confusion and the residual in the sky subtraction, these effects can lead to a mis-assignment of the redshift. To mimic this effect, a version has been generated with 1.5% catastrophic redshifts. These 1.5% of objects are randomly chosen and are assigned a new redshift following a uniform distribution. The redshift distribution has an impact on the velocity along the LOS direction and resembles the FOG effect as can be observed in Fig. 4.10. In the case of no smearing, we can only observe a large-scale squashing effect in Fig 4.10a. In the presence of redshift randomization, Fig. 4.10b and Fig. 4.10c) show a strong elongated feature that mimic the FoG feature. The mocks were originally built in a cubic box and then transformed into the sky coordinates with RA, DEC and Z. For the coordinate transformation, an observer is placed outside of the box, facing one of the 6 box sides. The chosen box side rotates among each of the 100 realizations of each HOD in a fixed order. Averaging the box sides is important for reducing the statistical error of the mocks (see discussion in section 5.3.1 in [Smith et al. \(2020\)](#)).

In order to test the flexibility of the model, we include a wide range of HOD configurations, some of which are even *not* motivated by the quasar physics. In the mocks we have tested so far, we only consider the halo clustering is determined by the halo mass. Some studies show that other halo properties may also contribute to the galaxy-halo connection. The environment-dependent halo formation time first proposed in [Sheth & Tormen \(2004\)](#) and then measured by [Gao et al. \(2005\)](#) demonstrated the impact of the halo formation time on the halo bias. Later on [Wechsler et al. \(2006\)](#) showed that the concentration can also affect the halo bias. These biases that are induced by the halo environments are called assembly bias and are extended as a function of halo spin, shape, mass accretion rate, and so on. So

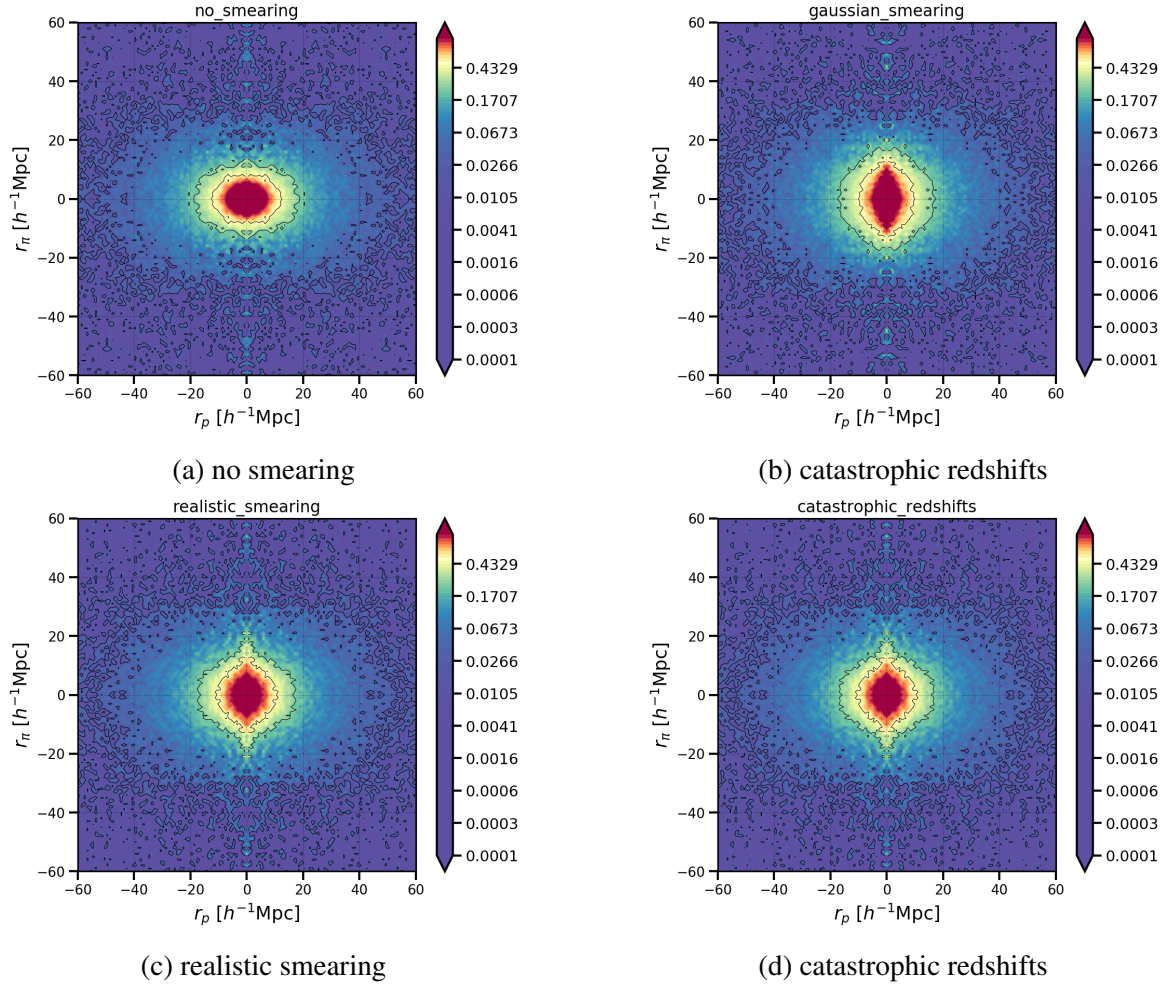


Fig. 4.10 Contours of the two-dimensional correlation function of the mock catalogues with different redshift distributions. The redshift smearing mimic a FOG effect that leads to a more pronounced elongated feature.

far, it is not clear how to fully establish a correlation between the halo properties and the observed clustering, and at the same time, some of the environment dependent effects can be fully absorbed into the clustering bias. For example, [Kobayashi et al. \(2019\)](#) showed for a set of BOSS-like mocks that, if one considers halo concentration induced assembly bias, the assembly bias leads to a boost in the galaxy clustering and leaves no impact on the RSD.

In our OUTERRIM mock challenge, we do not explicitly include the tests such as assembly bias, star formation rate, but their impacts should partially degenerate with the wide range of the HOD models available in our mock challenge test. A summary of the HOD models used in the non-blind mock challenge can be found in Table B.1.

We use the analytical method described in Section 4.3.2 to estimate the covariance matrix. The power spectrum is directly calculated from each configurations, and we use

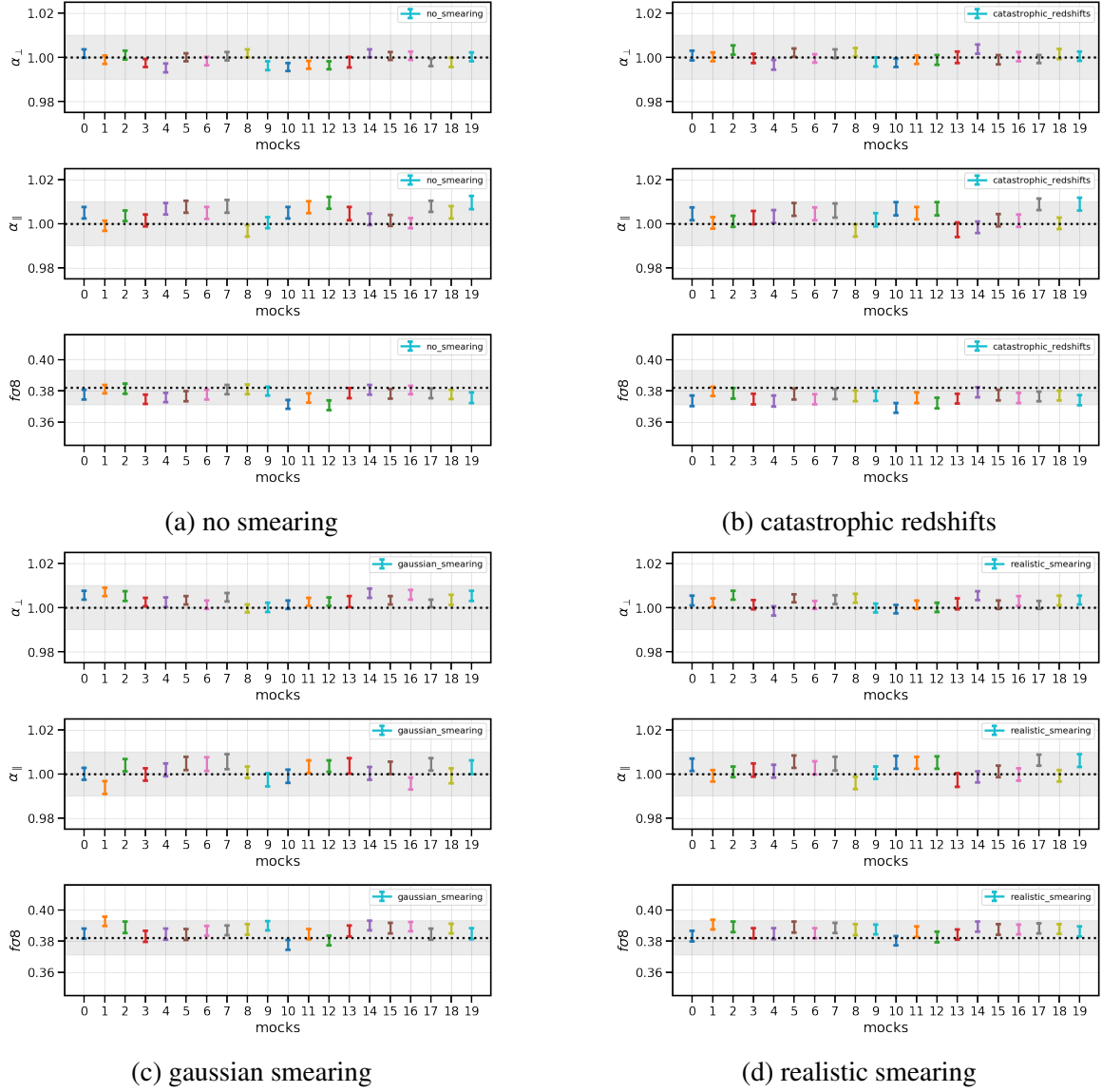


Fig. 4.11 Constraints on  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$  and  $f\sigma_8$  inferred from four redshift distribution schemes. Each panel shows the mean value of the fitted parameters as a function of the 20 HODs. The error bar for each colored points are the standard error of the mean (SEM). The grey band shows the 1% error for the AP parameters and 3% for  $f\sigma_8$ .

Eq. (4.31) to estimate the effective volume. We fit between the range  $s = [20, 160] h^{-1} \text{Mpc}$ , with bin separation  $ds = 8 h^{-1} \text{Mpc}$ . The fitting parameters and the priors can be found in Table 5.1. Additionally, we also compare the difference using  $ds = 5 h^{-1} \text{Mpc}$  for the first 5 mocks implemented with catastrophic redshift failure and found a root-mean-square (rms) of  $\delta\alpha_{\perp} = 0.002$ ,  $\delta\alpha_{\parallel} = 0.002$ , and  $\delta f\sigma_8 = 0.003$ . Fig 4.11 shows the fitting results to the OUTERRIM mocks. Each panel shows a different redshift randomization scheme and each

colored dot represents the mean fitted value out of the 100 realizations with the error bar representing the standard error of the mean (SEM). The grey bands give the 1% error in the AP parameters and 3% for  $f\sigma_8$ . Table 4.2 summarizes the statistics we obtain with the OUTERRIM nonblind mocks. Firstly, we are able to recover 1% accuracy for both the  $\alpha_{\parallel}$ ,  $\alpha_{\perp}$  and 3% for  $f\sigma_8$ . Secondly, we also observe a  $-3\%$  shift in  $f\sigma_8$  comparing the realistic smearing cases and the sample with 1.5% catastrophic redshift failure objects. We understand this as, the catastrophic redshift fully randomize the redshift and partly remove the information in the structure growth. Although the exact choice of the exact HOD formalism does not have a strong impact on the geometrical parameter neither the growth rate, it has a strong influence on the nuisance parameters, especially the one appeared the finger-of-god factor and the redshift uncertainty parameters such as  $a_{\text{vir}}$  and  $\sigma_{\text{zerr}}$ . Fig. 4.12 shows a clear dependence of the redshift uncertainty parameter on the satellite fraction. The random motions of the satellites change the quadrupole shape on scales  $\sim 40 h^{-1}\text{Mpc}$ , which leads to such strong dependency seen by the RSD-sensitive parameters. The linear relation exists not only for the ‘‘Gaussian smearing’’ case but also other redshift smearing schemes. Both the satellite fraction and redshift smearing effectively increase the redshift uncertainties and hence have an impact on the anisotropic clustering, but the redshift smearing can cause changes on even larger scales up to  $\sim 90 h^{-1}\text{Mpc}$ .

The systematic error is quantified from the mocks by taking the root-mean-square (rms) of the difference to the true cosmology. Using the mocks with realistic redshift smearing and catastrophic redshifts failures, we arrive at modelling systematics of  $\delta\alpha_{\perp} = 0.003$ ,  $\delta\alpha_{\parallel} = 0.004$ , and  $\delta f\sigma_8 = 0.008$ , as given in Table 4.4. Expressing the growth rate measurements have been expressed in terms of  $f\sigma_8$  has been a traditional choice. As explained in Section 3.3.2, [Sánchez \(2020\)](#) showed that the growth rate measurements can be better expressed in terms of  $f\sigma_{12}$ , so that the unit of the growth rate can be expressed in terms of quantity that is independent of the assumption of the cosmological parameter  $h$ , and we have  $\delta f\sigma_{12} = 0.007$ . Although this choice has no impact on the nonblind mock challenge, as we know precisely the true value of the cosmological parameters, this will be important for the blind mock challenge.

#### 4.4.2 Fiducial cosmology systematics<sup>2</sup>: blind mock challenge

To test the full analysis pipeline, we go one step further by testing our model ‘‘blindly’’. Since the OUTERRIM simulation is in a known cosmology, and it is computationally expensive to

<sup>2</sup>Fiducial cosmology here refer to both the set of cosmological parameters for the coordinates transformation and the ones for the generation of the template for the two-point correlation function. We do not distinguish the terminology because we always keep the same set of cosmological parameters for both.

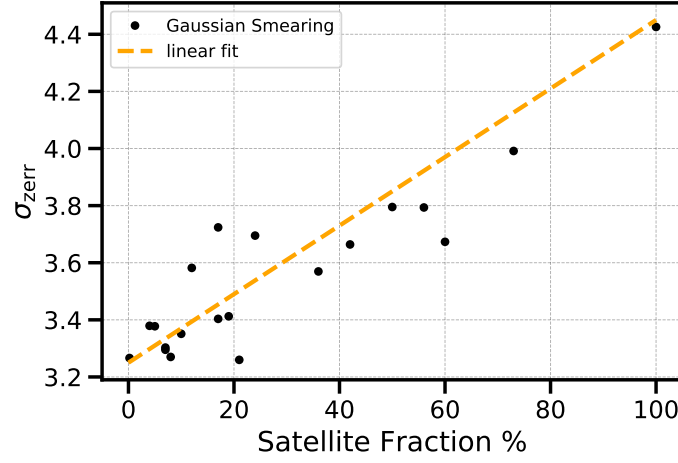


Fig. 4.12 Redshift uncertainty parameter as a function of the satellite fraction.

Table 4.2 Table summarizes the statistics we find with OUTERRIM nonblind mock test, the error in the table is the standard deviation mean of 100 mock OUTERRIM mocks for each configuration.

<b>no smearing (ns)</b>	<b>mean</b>	<b>rms</b>
$\alpha_{\perp}$	$0.998 \pm 0.019$	0.003
$\alpha_{\parallel}$	$1.003 \pm 0.026$	0.005
$f\sigma_8$	$0.377 \pm 0.030$	0.006
<b>catastrophic failure (cf)</b>	<b>mean</b>	<b>rms</b>
$\alpha_{\perp}$	$1.000 \pm 0.021$	0.003
$\alpha_{\parallel}$	$1.001 \pm 0.029$	0.004
$f\sigma_8$	$0.375 \pm 0.033$	0.008
<b>gaussian smearing (gs)</b>	<b>mean</b>	<b>rms</b>
$\alpha_{\perp}$	$1.003 \pm 0.020$	0.004
$\alpha_{\parallel}$	$1.000 \pm 0.029$	0.004
$f\sigma_8$	$0.385 \pm 0.033$	0.005
<b>realistic smearing (rs)</b>	<b>mean</b>	<b>rms</b>
$\alpha_{\perp}$	$1.001 \pm 0.020$	0.003
$\alpha_{\parallel}$	$1.001 \pm 0.028$	0.004
$f\sigma_8$	$0.387 \pm 0.033$	0.006



run a series of N-body simulation with the volume and resolution that meet the requirement of quasars analysis, we seek for an alternative method to rescale the OUTERRIM halo catalogue that mimics a new cosmology. The blind mock challenge uses the method described by [Mead & Peacock \(2014a,b\)](#). The idea is to rescale the halo comoving position vector and velocity to the one corresponding to a new cosmology  $\mathbf{x} \rightarrow \tilde{\mathbf{x}}$  and  $\mathbf{v} \rightarrow \tilde{\mathbf{v}}$ . The rescaling parameter  $s$  is found by minimizing the halo mass function, which is found to be nearly universal in form and depend on cosmology almost entirely through the linear variance of the density field  $\sigma$ ,

$$\delta_{\text{rms}}^2(s, z) = \frac{1}{\ln(R'_2/R'_1)} \int_{R'_1}^{R'_2} \frac{dR}{R} \left[ 1 - \frac{\sigma(R/s, z)}{\sigma'(R, z')} \right]^2, \quad (4.36)$$

where the prime  $\prime$  denotes the quantities in the target cosmology.  $R'_1$  and  $R'_2$  are the radius of the halos in the target cosmology that corresponds to the least and most massive halos in the original catalogues. The exact expression for linear variance  $\sigma^2$  with a top-hat of radius  $R$  at redshift  $z$  is expressed as equation (6) in [Mead & Peacock \(2014a\)](#). From the point of view of the halo model ([Cooray & Sheth, 2002](#); [Peacock & Smith, 2000](#), e.g.), the matching of the halo mass function guarantee the one-halo term that dominates the scales smaller than the virial radii of halos. In order to match the two-halo term, which is mostly responsible for the scales much larger than the virial radii that reproduces the large scale clustering, one needs to take the linear displacement field  $\psi$  in the rescaled cosmology and perturb the halo position using Zel'dovich approximation. The mapping of the position read

$$\tilde{\mathbf{x}} = \mathbf{x}' + \Delta \mathbf{x} \quad (4.37)$$

$$= s\mathbf{x} + b'(M') \left[ \sqrt{\frac{\Delta_{\text{lin}}'^2(k', z')}{\Delta_{\text{lin}}^2(sk', z)}} - 1 \right] \psi'_{k'}, \quad (4.38)$$

where the first term originates from matching the halo mass function and the second term comes from the displacement field. The bias  $b'(M')$  is directly measured from the catalogues, binned in mass and  $\Delta_{\text{lin}}$  is the dimensionless linear power spectrum. The linear displacement field  $\psi_k$  in k-space can be estimated from the density field  $\psi_k = -i \frac{\delta_k}{k^2} k$ . The natural measurement of the velocity  $U \equiv v/HaL$  is expected to be unaffected by the rescaling, except for the growth rate  $f$ . Such simple rescaling, although not perfect, can still recover the velocities on small scales by  $\sim 7\%$  ([Mead & Peacock, 2014a](#)). Given the relation between the displacement field and the velocity, the velocity should receive further correction by the Zel'dovich approximation due to the modification to the displacement field on the large scale. Similar to the rescaling of the position, the velocity can also be written as the composition of

two terms

$$\tilde{\mathbf{v}} = \mathbf{v}' + \Delta \mathbf{v} \quad (4.39)$$

$$= s \frac{(1+z)H'(z')f'(z')}{(1+z')H(z)f(z)} \mathbf{v} + \left[ \sqrt{\frac{\Delta_{\text{lin}}'^2(k'.z')}{\Delta_{\text{lin}}^2(sk',z)}} - 1 \right] \psi'_{k'}. \quad (4.40)$$

We produced in total 8 different cosmologies, with 3 HOD configurations for each. The choice of the cosmological parameters are given in Table B.2. The validation of the rescaling method can be found in [Smith et al. \(2020\)](#), which justifies the parameter range being tested in this work. To test the flexibility of the model, we let the  $\sigma_{\text{zerr}}$  parameter free during the blind analysis. Afterwards, we compare the difference in free varying or fixing the redshift error parameter  $\sigma_{\text{zerr}}$ . The difference we found is at sub-percent level for the inferred parameters, but the results tends to be slightly less biased without including the redshift error parameter  $\sigma_{\text{zerr}}$  when the mocks does not include the redshift smearing effect.

From the nonblind mock challenge, we already understood that our results are robust against different HOD. To further understand the impact purely due to the cosmology, we specifically analysed the rescaled mocks based on a set of fixed HODs, Table 4.3 summarizes the results in the fits for the blind mocks with fixed-HODs. We find that the inferred parameters are sensitive to the choice of the fiducial cosmology. We firstly notice a dependency in  $\Omega_{\text{m}}h^2$ . In the case of cosmo2, where the  $\Omega_{\text{m}}h^2$  is off by  $\sim 4\%$ , we find a deviation in  $\alpha_{\parallel}$  by  $\sim 2.4\%$ . Secondly, the parameters also show a dependence on the spectral index  $n_s$ . In the case of cosmo7, where the  $n_s$  is different by  $\sim 6.7\%$ , we have a shift in  $f\sigma_{12}$  for  $\sim 5\%$ . For our final results, we decide to add the effect of an incorrect fiducial cosmology as additional source of systematic error. To calculate a systematic error due to the fiducial cosmology, we calculate the rms of the set of 24 blind mocks, which are then added in quadrature to the modelling systematic error calculated from the non-blind mocks. Table 4.4 lists our results from the “blind” mock challenge, and the last column is calibrated w.r.t the offset in the “no smearing” non-blind mocks with free  $\sigma_{\text{zerr}}$ . The rms we find with the blind mocks challenge is  $\delta\alpha_{\perp} = 0.007$ ,  $\delta\alpha_{\parallel} = 0.011$ , and  $\delta f\sigma_8 = 0.011$ . When expressing in terms of  $f\sigma_{12}$ , we have  $\delta f\sigma_{12} = 0.009$ .

#### 4.4.3 Observational systematics: EZMOCKS

We use the EZMOCKS to test the observational systematics. EZMOCKS ([Chuang et al., 2015](#)) stands for effective Zel’dovich mock, where the initial condition is set by solving the perturbation theory with Zel’dovich approximation. After obtaining the dark matter



Table 4.3 Shifts in the parameters obtained from the OUTERRIM blind mock challenge with fixed HODs. Each row represents a different box with different combination of cosmology parameters.

	$\delta\alpha_{\perp}$	$\delta\alpha_{\parallel}$	$\delta f\sigma_{12}$	$\delta f\sigma_8$
cosmo0	$0.0064 \pm 0.0020$	$-0.0002 \pm 0.0027$	$0.0030 \pm 0.0029$	$-0.0014 \pm 0.0030$
cosmo1	$0.0142 \pm 0.0022$	$0.0094 \pm 0.0022$	$0.0040 \pm 0.0025$	$0.0088 \pm 0.0026$
cosmo2	$0.0079 \pm 0.0033$	$0.0238 \pm 0.0041$	$0.0061 \pm 0.0032$	$-0.0042 \pm 0.0033$
cosmo3	$0.0114 \pm 0.0017$	$0.0030 \pm 0.0025$	$0.0065 \pm 0.0028$	$0.0146 \pm 0.0029$
cosmo4	$-0.0000 \pm 0.0020$	$-0.0001 \pm 0.0029$	$-0.0014 \pm 0.0027$	$0.0036 \pm 0.0028$
cosmo5	$0.0010 \pm 0.0017$	$-0.0038 \pm 0.0022$	$-0.0049 \pm 0.0022$	$-0.0084 \pm 0.0023$
cosmo6	$0.0054 \pm 0.0019$	$-0.0003 \pm 0.0024$	$0.0122 \pm 0.0028$	$0.0063 \pm 0.0030$
cosmo7	$-0.0093 \pm 0.0017$	$-0.0109 \pm 0.0020$	$-0.0191 \pm 0.0031$	$-0.0238 \pm 0.0032$

Table 4.4 root-mean-square (RMS) on the blind and nonblind mocks test. For blind mock challenge, in total 24 mocks with 8 different cosmologies are used for the statistics. Each cosmology is repeated 3 times with different HOD settings. For nonblind mock challenge, in total 20 mocks were used.

Blind	$\delta\alpha_{\perp}$	$\delta\alpha_{\parallel}$	$\delta f\sigma_{12}$	$\delta f\sigma_8$
	0.007	0.011		
	$\delta_{D_M/r_{\text{drag}}}$	$D_H/r_{\text{drag}}$	0.009	0.011
	0.209	0.145		
Non-blind	$\delta\alpha_{\perp}$	$\delta\alpha_{\parallel}$	$\delta f\sigma_{12}$	$\delta f\sigma_8$
	0.003	0.004		
	$\delta_{D_M/r_{\text{drag}}}$	$D_H/r_{\text{drag}}$	0.007	0.008
	0.070	0.057		

field, the halos were obtained by the effective bias model. Certain scatters are introduced to the identification of the spherically collapsing structure to account for the stochastic and non-local bias. The probability density function (PDF) of the dark matter haloes is calibrated by mapping the density field to the BigMulti-Dark (BigMD) N-body simulations (Klypin et al., 2016). Finally, the galaxies are assigned to the dark matter particles. The EZmock boxes were generated at different redshift snapshot. In the case of quasars, 7 redshift shells were applied, each mock in the shell has the same initial Gaussian density field. The light-cone was constructed out of the 7 redshift shell, which also include the effects for RSD, footprint and radial selection. The rest of the EZMOCKS parameters are calibrated to the data, independently for NGC and SGC. In this way, the redshift error for the quasars

intrinsically encoded in the EZMOCKS, as it was calibrated with respect to the real data. The EZMOCKS have flat  $\Lambda$ CDM cosmology, with matter density parameter  $\Omega_m = 0.307$ , baryon density  $\Omega_b h^2 = 0.022$ , a dimensionless Hubble parameter  $h = 0.678$ , and no contribution from massive neutrinos. The power spectrum of these mocks is characterized by a scalar spectral index  $n_s = 0.96$ , normalized to a value of  $\sigma_8(z = 0) = 0.8225$ . These parameters correspond to a value of  $f\sigma_8(z = 1.52) = 0.378$  at the mean redshift of the quasar sample.

We consider the impact of the following observational effects: spectroscopic redshift failures, close pairs, and the photometric calibration. We start with the mocks including the angular selection function. In the first step, contamination from the data were added to the mocks, which includes stars, redshift failure objects, wrong objects classes, the objects have no chances to receive a good redshifts (e.g. unplugged fibers, etc.) and the not tiled objects. For the photometric correction, a fit is applied to the minimize the trend in the ratio  $n_{\text{QSO}}/n_{\text{random}}$  in the extinction corrected g-band depth and the stellar density. The quasar sample has a special Legacy group, these objects do not receive any spectroscopic or collision correction and are separately assigned with tag  $\text{IMATCH} = 2$  to distinguish from the CORE sample. Due to the finite radius of the fiber cladding, it is inevitable when two objects are angularly close to each other, only one can be assigned with a fiber, which is referred to the fiber collision. Such effect is added to the EZMOCKS using the FiberCollision module from the NBODYKIT (Hand et al., 2018). After identifying the close pairs, the objects fall in the same collision groups are up-weight be the ratio  $n_{\text{hasfiber}}/n_{\text{total}}$ . The redshift failure corrections are applied using the same definition as in Eq. (4.4), and the mock objects inherit the property from the data objects through closest angular-matching. The code for post-processing the systematic effects on the mocks is integrated into the clustering analysis toolkit.<sup>3</sup>

Fig. 4.13 compares the impact of the different systematics on the multipoles. It can be seen that the largest effect on small scales is due to fibre collisions (orange curve). For the monopole, the impact is visible from scales  $s \lesssim 25 h^{-1}\text{Mpc}$ . For higher order multipoles, this effect is already visible at scales starting from  $s \lesssim 50 h^{-1}\text{Mpc}$ .

There are several methods that can be utilized to correct the small-scale clustering measurements for the effect of fibre collisions. This includes an angular up-weighting (e.g. Hawkins et al., 2003), modelling the effect of fibre collisions on the correlation function (Hahn et al., 2017), or an inverse pair weighting scheme (e.g. Bianchi & Percival, 2017). To assess the systematics due to the fibre collision, it is required that the radial distribution of the ‘unobserved’ objects is similar to the one of the total objects. It is not necessary that the collided objects which are identified within the same group are physically associated.

<sup>3</sup>[https://github.com/julianbautista/eboss\\_clustering](https://github.com/julianbautista/eboss_clustering)

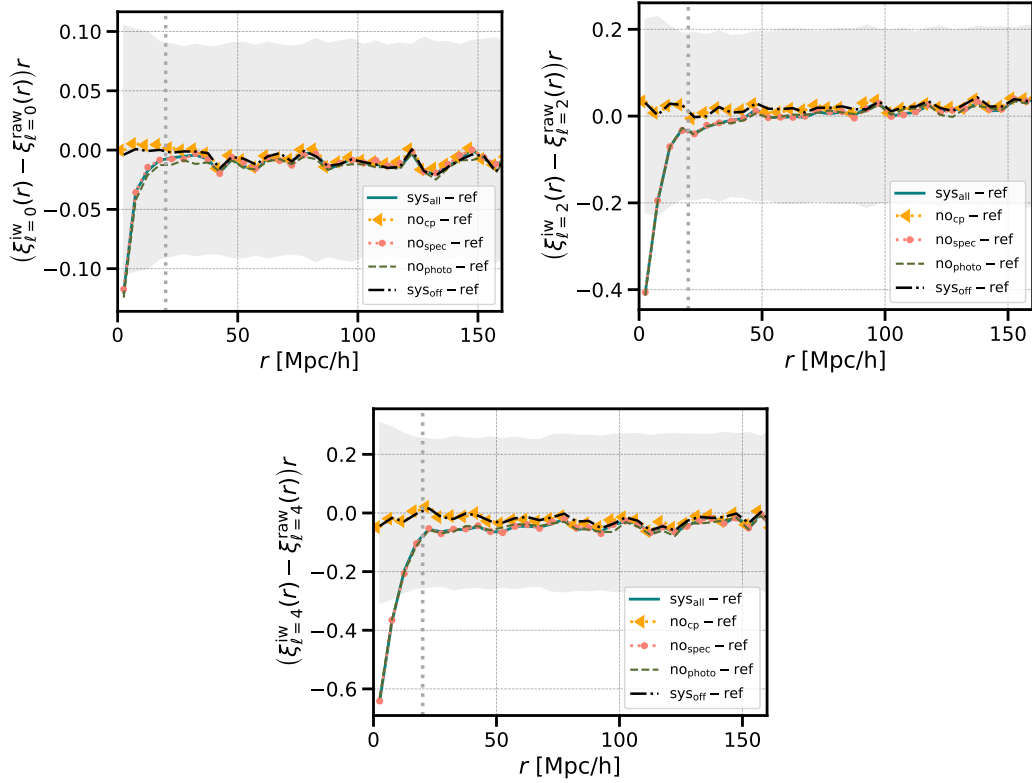


Fig. 4.13 Comparison of the multipoles for EZMOCKS with different combination of systematics with respect to the reference one, which does not include any systematics or radial integral constraints effect.

Therefore, it is not critical whether EZMOCKS predicts as accurate small scale clustering as the N-body simulations. Fig. 4.14 shows the radial distribution of the unobserved objects and the total objects in one of the EZMOCKS realizations (left panel), as well as their ratio as a function of redshift (right panel). The similarity of the radial distribution between the unobserved and the total objects in the post-processed mocks making it viable to use these mocks for assessing the systematics.

To correct for the effect of fibre collisions, the method we use is based on [Hahn et al. \(2017\)](#), which models the effect of fibre collisions on the correlation function. This method produces similar results to a more rigorous pair weighting scheme (see Section 5.2). The effect of fibre collisions is treated as a top hat function in configuration space. Since our model is built in Fourier space, it is more convenient to modify the power spectrum directly by convolving it with the Fourier transform of the top hat function. However, this leads to the technical difficulties of integrating  $k$  to infinity, and evaluating high order multipoles (up to  $\ell = 18$ , as in the original paper). Therefore, we have implemented the method both in

configuration and Fourier space, and have verified that the difference between the two is very small.

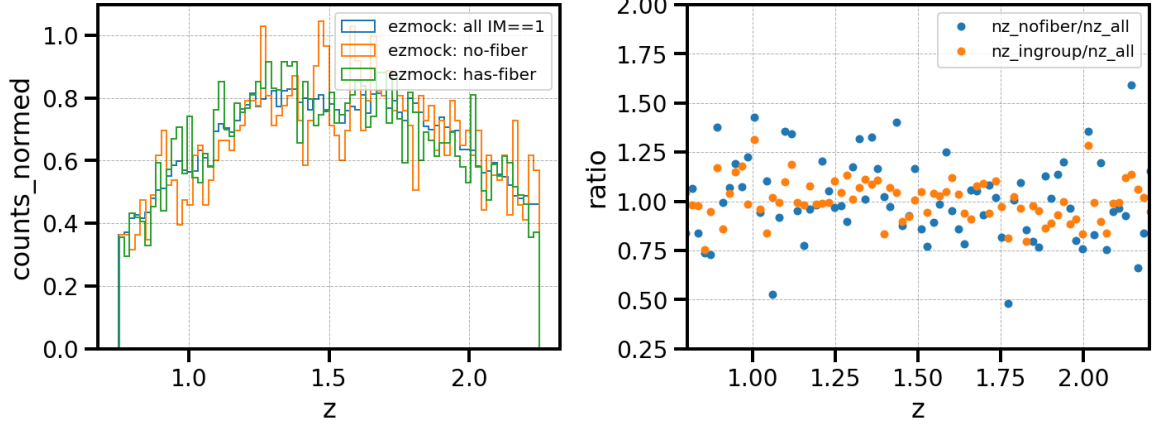


Fig. 4.14 Left panel: distribution of the  $n(z)$  in one realization of the EZMOCKS. The blue histogram is the distribution of the total objects. The green histogram is the distribution for those objects that are assigned with a fibre, and the orange one corresponds to those do not receive a fibre assignment. Right panel: blue dots denote the ratio of  $n(z)$  between the objects assigned (orange) or not assigned (blue) with a fiber and the total objects. The average difference between the total and “unobserved” objects in the radial distribution is less than 1%

The projected correlation function measured from the EZMOCKS on small scales is shown in the left panel of Fig. 4.15. For the full mock, with no fibre collisions ( $w_p^{\text{true}}$ ), the clustering amplitude is approximately zero. This is because pairs of physically associated quasars at these separations are very rare, and most of the pairs are due to random alignments on the sky.  $w_p^{\text{NN}}$  indicates the clustering measured from the mocks with fibre collisions, that has been corrected with a nearest neighbour (NN) weight. The negative clustering amplitude indicates an ‘anti-correlation’ due to the fibre collision, but  $w_p^{\text{NN}}$  does not reach  $-1$ , since a fraction of closely separated pairs can still be observed, due to the overlapping regimes and the Legacy objects.

The right panel of Fig. 4.15 shows the ratio of the two projected correlation functions. This function is sloped between  $0.5 \lesssim r_p \lesssim 1.0 h^{-1} \text{Mpc}$ , making a top hat function a poor fit. This is because the fibre collision scale corresponds to a physical scale that depends on redshift, varying from  $D_{\text{fc}}(z_{\text{min}}) = 0.58 h^{-1} \text{Mpc}$  to  $D_{\text{fc}}(z_{\text{max}}) = 1.13 h^{-1} \text{Mpc}$ .

Starting from equation (23) in (Hahn et al., 2017), the correction can be written in terms of the configuration space multipoles,

$$\Delta \xi_\ell = -f_s(2\ell + 1) \int_0^1 W_{\text{fc}} \left( s\sqrt{1 - \mu^2} \right) (\xi(s, \mu) + 1) \mathcal{L}_\ell(\mu) d\mu, \quad (4.41)$$

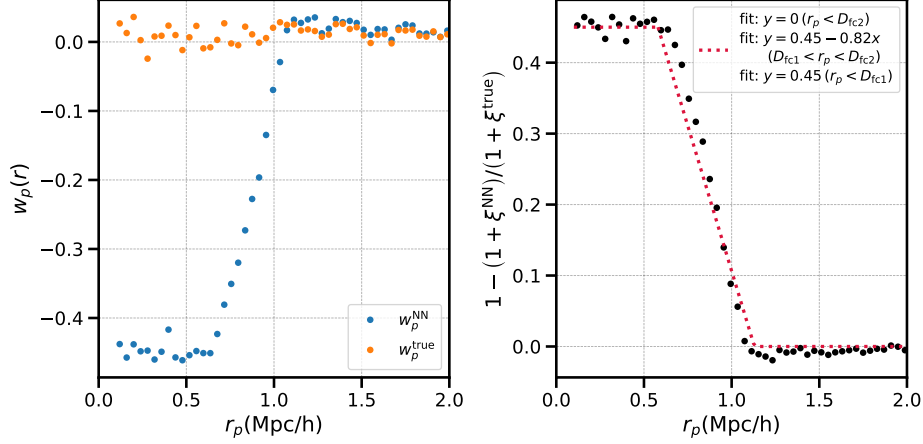


Fig. 4.15 Left panel: projected correlation function measured from the EZMOCKS with no fibre collisions ( $w_p^{\text{true}}$ , orange), and with fibre collisions and a nearest neighbour weighting ( $w_p^{\text{NN}}$ , blue). Right panel: the ratio  $1 - (1 + \xi^{\text{NN}})/(1 + \xi^{\text{true}})$  (black dots), with a best fit model (dotted red curve). The turnover scales are  $D_{\text{fc1}} = 0.58 h^{-1} \text{Mpc}$  and  $D_{\text{fc2}} = 1.13 h^{-1} \text{Mpc}$ .

We use two different functional forms for  $W_{\text{fc}}(x)$ . The first function we use is the original top hat function, where the step is at the scale  $r_p = D_{\text{fc}}(z_{\text{eff}}) = 0.91 h^{-1} \text{Mpc}$ . It is natural to introduce a cut in the line of sight direction, with  $\mu_c = \sqrt{1 - \frac{r_p^2}{s^2}}$ , and therefore Eq. (4.42) can be simplified to

$$\Delta \xi_\ell = -f_s(2\ell + 1) \left[ \int_{\mu_c}^1 (1 + \xi(s, \mu)) \mathcal{L}_\ell(\mu) d\mu \right], \quad (4.42)$$

We also use a functional form for  $W_{\text{fc}}(x)$  that is motivated by Fig. 4.15, which we define as

$$W_{\text{fc}}(r_p) = \begin{cases} 1 & \text{for } r_p \leq D_{\text{fc1}} \\ t_p - k_p r_p & \text{for } D_{\text{fc2}} \geq r_p > D_{\text{fc1}} \\ 0 & \text{for } r_p > D_{\text{fc2}} \end{cases} \quad (4.43)$$

The slope  $k_p$  and intercept  $t_p$  are determined by the two characteristic scales,  $D_{\text{fc1}}$  and  $D_{\text{fc2}}$  as well as the fraction of non-overlapping area,  $f_s$ , which we leave as a free fitting parameter.

We divide them into two groups: in the first group, we exam the effect associated to the radial integral constraint (RIC; [de Mattia & Ruhlmann-Kleider, 2019](#)). We used a set of mocks  $w_{\text{baseline}}$ , which are only downsampled by completeness and the redshifts for the random catalogue are drawn from a single global file. Then we added the RIC effect by drawing the redshifts for the random catalogues from each individual data mock  $w_{\text{baseline}}^{\text{ric}}$ . In the next line, we correct for this effect following [de Mattia & Ruhlmann-Kleider \(2019\)](#) and

denote it as  $w_{\text{baseline}}^{\text{ric-corr}}$ . In the second group, we exam the effects associated by the observational effects.  $w_{\text{no-sys}}$  are mocks without applying any systematics. Since fibre collisions have the largest impact on the correlation function among the observational systematics (Fig. 4.13), we show in Table 4.5 the results where all systematics are applied, including and excluding fibre collisions ( $w_{\text{all}}$  and  $w_{\text{nocp}}$ , respectively). We show results using the top hat,  $w_{\text{all}}^{\text{fc(top-hat)}}$ , or trapezoidal function,  $w_{\text{all}}^{\text{fc(trapezoid)}}$ , to apply a correction. When applying the trapezoidal correction, we initially left  $f_s$  as a free parameter, but found a best fit value of  $f_s = 0.45$  which coincides very well with the predicted value from Fig 4.15. Hereafter, we keep this parameter fixed. Uncertainties are the standard error of the mean from the 1000 EZMOCKS.

To estimate the final observational systematics, we add the RIC effect ( $\Delta^{\text{ric}}$ ) and the observational effects ( $\Delta^{\text{obs}}$ ) in quadrature. We quote the final systematics,  $\delta_{\text{sys}} = \max \{ \Delta_{\text{sys}}, 2\sigma_{\text{stat}} \}$ , as the larger value between the systematic bias and two times the standard error of the mean for the mocks, the  $2\sigma_{\text{stat}}$  and we arrive at  $\delta\alpha_{\perp} = 0.003$ ,  $\delta\alpha_{\parallel} = 0.005$ , and  $\delta f\sigma_8 = 0.004$ .

## 4.5 BAO-only analysis

In addition to the full-shape analysis, we also present BAO-only measurements of the geometric parameters  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  as an additional consistency check. These measurements attempt to isolate the BAO information such that none of the constraining power comes from information in the broad-band amplitude of the correlation function. We follow the same methodology as in [Ross et al. \(2017\)](#), which was itself based on [Xu et al. \(2013\)](#) and [Anderson et al. \(2014\)](#). The BAO feature is isolated in Fourier-space and damped as a function of  $\mu$  in order to approximate the effects of non-linear structure formation and redshift-space distortions

$$P_{\text{BAO}}(k, \mu) = (P_{\text{lin}} - P_{\text{nw}}) e^{-k^2 \cdot \sigma_s(\mu)^2} + P_{\text{nw}}. \quad (4.44)$$

The linear power spectrum  $P_{\text{lin}}$  is obtained from CAMB, while the “no-wiggle” power spectrum is from the fitting formulae by [Eisenstein & Hu \(1998\)](#). The factor in the exponential term that captures the damping of the BAO feature under the nonlinear evolution is given by,

$$\sigma_s^2 = (1 - \mu^2) \Sigma_{\perp}^2 / 2 + \mu^2 \Sigma_{\parallel}^2 / 2, \quad (4.45)$$

with  $\Sigma_{\perp} = 3h^{-1}\text{Mpc}$ ,  $\Sigma_{\parallel} = 8^{-1}\text{Mpc}$ . These parameters match those adopted by [Neveux et al. \(2020\)](#) for the Fourier-space analysis.

Table 4.5 Observational systematics measured from the EZMOCKS, showing the offsets in the measured values of  $\alpha_\perp$ ,  $\alpha_\parallel$  and  $f\sigma_8$  to the fiducial values, with different systematics applied. The offset is inferred from the fits to the mean of the 1000 correlation function multipoles. The first group shows the effect associated to the radial integral constraint (RIC; [de Mattia & Ruhlmann-Kleider, 2019](#)). Mocks  $w_{\text{baseline}}$  are only downsampled by completeness and the redshifts for the random catalogue are drawn from a single global file.  $w_{\text{baseline}}^{\text{ric}}$  are the same, but redshifts in the random catalogues are drawn from the data mocks. The RIC effects are corrected in the model for  $w_{\text{baseline}}^{\text{ric-corr}}$ . The second group shows the effects related to the observational effects.  $w_{\text{no-sys}}$  are mocks without including observational systematics.  $w_{\text{nocp}}$  includes all systematics except for fibre collisions, while  $w_{\text{all}}$  includes all systematics. The next rows show the result after applying the correction of [Hahn et al. \(2017\)](#), using a top hat function and a trapezoidal function. Uncertainties are taken from the standard error of the mean of the 1000 EZMOCKS.

systematics	$\Delta\alpha_\perp$	$\Delta\alpha_\parallel$	$\Delta f\sigma_{12}$
$w_{\text{baseline}}$	$0.002 \pm 0.001$	$-0.003 \pm 0.001$	$-0.009 \pm 0.001$
$w_{\text{baseline}}^{\text{ric}}$	$0.006 \pm 0.001$	$-0.005 \pm 0.001$	$-0.013 \pm 0.001$
$w_{\text{baseline}}^{\text{ric-corr}}$	$0.003 \pm 0.001$	$-0.004 \pm 0.001$	$-0.012 \pm 0.001$
$\Delta^{\text{ric}}$	$0.001 \pm 0.001$	$-0.001 \pm 0.001$	$-0.003 \pm 0.001$
$w_{\text{no-sys}}$	$0.009 \pm 0.001$	$0.002 \pm 0.002$	$-0.006 \pm 0.001$
$w_{\text{nocp}}$	$0.008 \pm 0.001$	$0.002 \pm 0.002$	$-0.006 \pm 0.001$
$w_{\text{all}}$	$0.017 \pm 0.001$	$-0.008 \pm 0.002$	$0.008 \pm 0.002$
$w_{\text{all}}^{\text{fc(top-hat)}}$	$0.011 \pm 0.001$	$0.002 \pm 0.002$	$-0.003 \pm 0.002$
$w_{\text{all}}^{\text{fc(trapezoid)}}$	$0.010 \pm 0.001$	$-0.001 \pm 0.002$	$-0.004 \pm 0.002$
$\Delta^{\text{obs}}$	$0.001 \pm 0.001$	$-0.003 \pm 0.002$	$0.002 \pm 0.002$
Total	0.003	0.005	0.004

In order to approximate the effects of redshift-space distortions on the broadband  $P(k)$ , we use,

$$P(k, \mu) = \left( \frac{1 + \mu^2 \beta}{1 + k^2 \mu^2 \Sigma_s^2 / 2} \right)^2 P_{\text{BAO}}(k). \quad (4.46)$$

Broad-band polynomial terms are included in the model in order to allow considerable freedom in the broadband, but the inclusion of the factor above allows the fiducial model to be in reasonable agreement before their inclusion. The factor  $\beta$  is fixed at 0.4; for a physical redshift-space distortion model this is ratio for the growth rate and the linear bias is given by  $\beta = f/b_1$ . Here, it controls the fiducial overall amplitude of the quadrupole, which is allowed to vary in the BAO fits through the  $B_0$  and  $B_2$  terms defined below. The term  $\Sigma_s$  relates to

the expected contribution from satellite velocities and redshift errors. We set  $\Sigma_s = 4^{-1} \text{Mpc}$ , again matching the choice adopted by [Neveux et al. \(2020\)](#).

The correlation function BAO template is then a Fourier transform of Eq. (4.46). The  $\alpha$  are applied to the template via

$$\xi_{\text{temp}}(s, \alpha_{\perp}, \alpha_{\parallel})_F = \int_0^1 d\mu F(\mu') \xi_{\text{temp}}(s', \mu'), \quad (4.47)$$

where  $F$  is an arbitrary window over  $\mu$  (defined for our particular case below),  $\mu' = \mu \alpha_{\parallel} / \sqrt{\mu^2 \alpha_{\parallel}^2 + (1 - \mu^2) \alpha_{\perp}^2}$ , and  $s' = s \sqrt{\mu^2 \alpha_{\parallel}^2 + (1 - \mu^2) \alpha_{\perp}^2}$ .

We fit for monopole  $\ell = 0$  and quadrupole  $\ell = 2$  and we have,

$$\xi_0^{\text{mod}}(s) = B_0 \xi_0^{\text{temp}}(s, \alpha_{\perp}, \alpha_{\parallel}) + A_0(s), \quad (4.48)$$

$$\xi_2^{\text{mod}}(s) = \frac{5}{2} \left( B_2 \xi_{\mu 2}^{\text{temp}}(s, \alpha_{\perp}, \alpha_{\parallel}) - B_0 \xi_0(s, \alpha_{\perp}, \alpha_{\parallel}) \right) + A_2(s), \quad (4.49)$$

where the polynomial  $A_x(s) = a_{x,1}/s^2 + a_{x,2}/s + a_{x,3}$  removes information from the broad-band shapes of the  $\xi_{\ell}$ ,  $B_x$  adjusts the amplitude of the BAO feature, and we denote  $\xi_{\mu 2} \equiv 3 \int_0^1 d\mu \mu^2 \xi(\mu)$ . In order to obtain the likelihood for  $\alpha_{\parallel}, \alpha_{\perp}$ , we find the minimum  $\chi^2$  over a grid in their values in the range  $0.8 < \alpha_{\parallel} < 1.2$  and  $0.8 < \alpha_{\perp} < 1.2$ .

We will also obtain BAO results using only  $\xi_0$ . In this case we use the same modeling and nuisance parameters for  $\xi_0$ , except we simply have

$$\xi_0^{\text{temp}}(s, \alpha_{\text{iso}}) = \xi_0^{\text{temp}}(s \alpha_{\text{iso}}). \quad (4.50)$$

In the case where information is distributed in a spherically symmetric manner,  $\alpha_{\text{iso}} = \alpha_{\perp}^{2/3} \alpha_{\parallel}^{1/3}$  and this is the best constrained combination of the BAO information. We obtain the likelihood for  $\alpha_{\text{iso}}$  by finding the  $\chi_{\text{min}}^2(\alpha_{\text{iso}})$  on a grid  $0.8 < \alpha_{\text{iso}} < 1.2$ .



# Chapter 5

## Cosmological implication from eBOSS

### 5.1 Constraints from the eBOSS QSOs DR16 QSO analysis

In this section we explore the BAO and RSD constraints for DR16 QSO data in terms of comoving angular diameter distance, Hubble distance and the growth rate of the cosmic structure. We estimate the effective redshift of the eBOSS DR16 QSO sample using the definition,

$$z_{\text{eff}} = \frac{\sum_{i,j} w_i w_j (z_i + z_j)}{\sum_{i,j} 2w_i w_j}, \quad (5.1)$$

where we sum over pairs with a separation distance between  $20 h^{-1} \text{Mpc} \leq ds \leq 160 h^{-1} \text{Mpc}$ , the weights  $w_i$  are defined as in Eq. (4.5). The exact definition of the pair separation distance has marginal impact on the effective redshift. A comparison using different definition of effective redshift can be found in Section 5.2.1.

#### 5.1.1 Results in the configuration space: full-shape analysis

The final parameter inference is performed based on RESPRESSO + fitting function and the RSD model as described in Section 3.1.4 and Section 3.3.1. We use the same bias model as in DR14 analysis, but leave the nonlocal bias parameter  $\gamma_3^-$  as a free parameter. We include the correction for the fiber collision effect using the trapezoidal model, as a modification to the [Hahn et al. \(2017\)](#) method. We iteratively find the parameter  $f_s = 0.4$ , which agrees well with Fig. 5.8, which shows the projected correlation function for the data in the north cap (green dots), south cap (red dots) and the combined pair counts (grey), where the error bar is obtained from the mocks. We perform the analysis on the multipoles  $\xi_\ell = 0, 2, 4(s)$

within range  $20 h^{-1}\text{Mpc} \leq s \leq 160 h^{-1}\text{Mpc}$  for bin separation  $ds = 5 h^{-1}\text{Mpc}$  and use 1000 EZMOCKS, which include both photometric and spectroscopic systematics to estimate the covariance matrix. Fig 5.1 shows the posterior distribution of the AP parameters as well as the  $f\sigma_{12}$ , for NGC (orange), SGC (blue) and the combined in pair counts for both north and south caps (pink). Fig. 5.2 shows the error on the data for AP parameters and  $f\sigma_{12}$

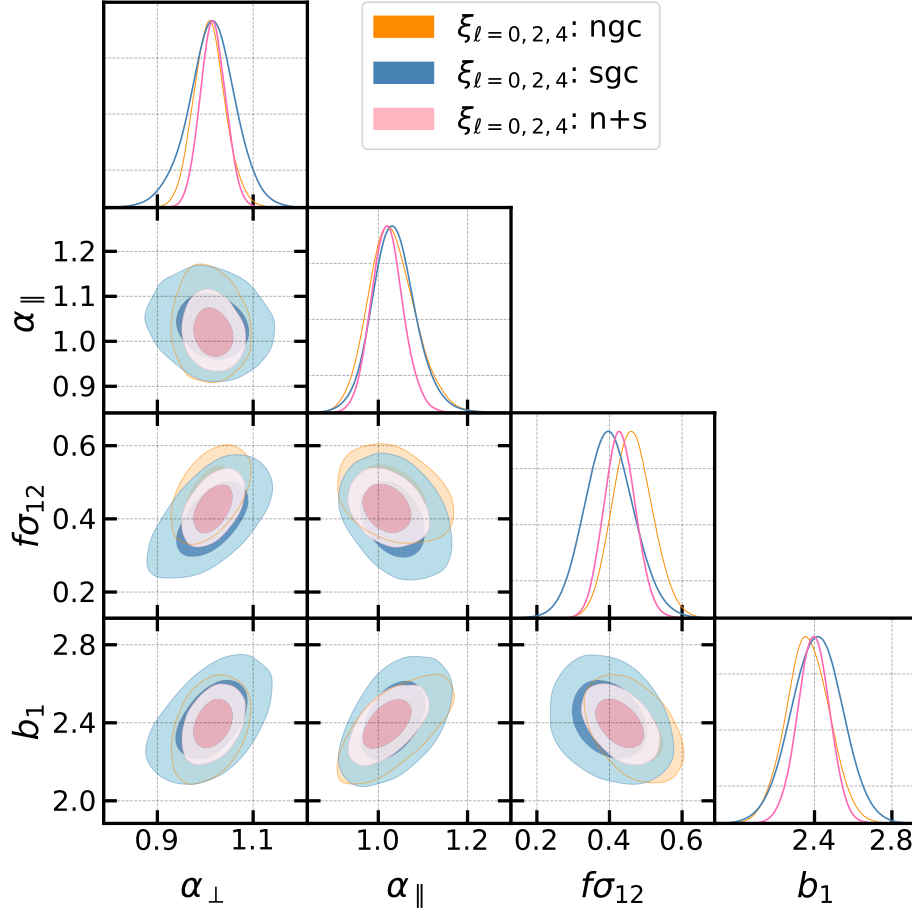


Fig. 5.1 Posterior distribution for AP parameters,  $f\sigma_{12}$  and linear bias  $b_1$  for NGC (orange), SGC (blue), and combined (pink).

compared to the distribution of the error from the mocks for the combined north and south caps. The error inferred from the data sits at the lower tail of the mocks. There are two reasons to account for this. Firstly, is the BAO signal in the real data is higher than that in the averaged mocks, and effectively we have a stronger SNR that helps to reduce the statistical error. Secondly, the ratio between the number of objects in the random catalogue to the data catalogue is 2.5 times higher for the data than that in the mocks, which leads to an effective noisier estimation of the error on the mocks.

Table 5.3 lists our measured values, in terms of the AP parameters. The error bars are derived statistically from the Monte Carlo Markov Chain (MCMC) chain with the correction factor  $\sqrt{M} = 1.036$  for the statistical error.

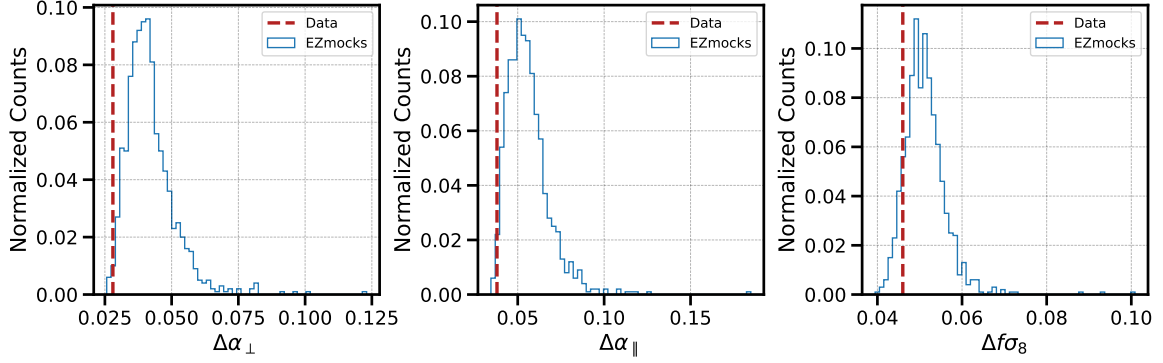


Fig. 5.2 Distribution of the statistical error for  $\Delta\alpha_{\perp}$ ,  $\Delta\alpha_{\parallel}$  and  $f\sigma_{12}$  on the EZMOCKS with all observational effects included (blue histogram) and the statistical error of the DR16 QSO data (red dashed line)

We adopt the same fiducial cosmology as for DR14 analysis,  $\{\Omega_m, \Omega_b h^2, h, n_s, \sigma_8\} = \{0.31, 0.022, 0.676, 0.97, 0.8\}$ , where the total matter density parameter also includes a contribution from massive neutrinos  $\sum_i m_\nu^i = 0.06\text{eV}$ , corresponding to  $\Omega_\nu h^2 = 0.0064$ , and  $\sigma_{12}$ . We obtain the fiducial distances,  $H(z = 1.48) = 157.40 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $D_M(z = 1.48) = 4446.82 \text{ Mpc}$ , and  $r_{\text{drag}} = 147.8 \text{ Mpc}$ . This corresponds to the amplitude of the matter power spectrum  $\sigma_{12} = 0.79$  in the units of Mpc (without  $h$ ), as suggested in [Sánchez \(2020\)](#). Finally, using Eq. (4.12) and (4.16), we arrive at the comoving angular diameter distance, Hubble distance, and  $f\sigma_{12}$ :

$$D_M(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.66(84)(22)(10), \quad (5.2)$$

$$D_H(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 13.11(49)(16)(6), \quad (5.3)$$

$$f\sigma_{12}(z_{\text{eff}} = 1.48) = 0.435(46)(11)(4), \quad (5.4)$$

In addition, we also have  $f\sigma_8$ :

$$f\sigma_8(z_{\text{eff}} = 1.48) = 0.439(46)(13)(4), \quad (5.5)$$

where the first parenthesis denotes the statistical error which includes correction factor  $\sqrt{M} = 1.036$ , the second parenthesis denotes the systematic error from the theoretical modelling and the fiducial template (summed in quadrature), and the last parenthesis is for the observational

systematics. These numbers are listed in Table 5.2. The systematic errors are quoted as the larger value between the bias and the  $2\sigma$  of the standard deviation of the mean of the mocks.

Fig. 5.3 shows the redshift evolution of the distance measurements (left panel) and growth rate measurement (right panel). Our final results from the DR16 quasar sample in the configuration space are shown by the yellow points with error bars. We compare this to the  $\Lambda$ CDM model inferred from the Planck CMB temperature and polarization measurements. We also show previous results from the SDSS main galaxy sample for the distance measurement (Ross et al., 2015a) and growth rate measurement (Howlett et al., 2015), the constraints from BOSS DR12 LRG sample (Alam et al., 2017), and the combined constraints from eBOSS DR14 Ly $\alpha$  measurements (Blomqvist et al., 2019; de Sainte Agathe et al., 2019). The orange error bars are from this work using the quasar sample in configuration space. With the final sample, statistically we gain  $\sim 45\%$  in the distance measurement, and  $\sim 30\%$  in the growth rate measurement compared to our DR14 analysis.

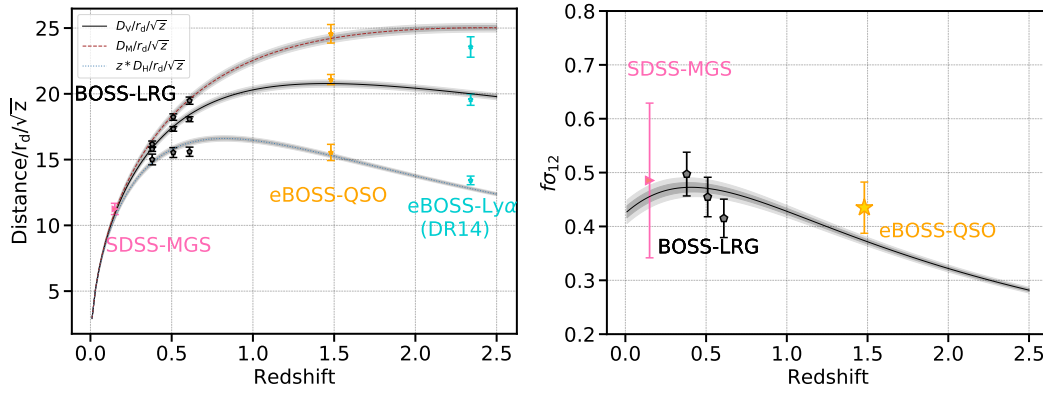


Fig. 5.3 Redshift evolution of the distance parameter and the cosmic growth rate. The  $\Lambda$ CDM model, with input from the Planck 2018 MCMC chains, is shown by the curves (as indicated in the legend). The grey points are from the BOSS DR12 LRG sample (Alam et al., 2017), the blue points are from the combined eBOSS DR14 Ly $\alpha$  auto- and cross-correlation function (Blomqvist et al., 2019; de Sainte Agathe et al., 2019), the pink points are from an early SDSS MGS sample (from Ross et al. (2015a) for the distance measurement and from Howlett et al. (2015) for the growth rate). The orange points show the final results from the eBOSS DR16 quasar analysis, in configuration space.

We present the parameter covariance matrix including the statistical error, theoretical modelling systematics and observational systematics in the  $D_M/r_{\text{drag}}$ ,  $D_H/r_{\text{drag}}$ , and  $f\sigma_{12}$

basis as

$$\mathbf{C}_\lambda = \begin{pmatrix} D_M/r_{\text{drag}} & D_H/r_{\text{drag}} & f\sigma_{12} \\ 7.709 \times 10^{-1} & -5.656 \times 10^{-2} & 1.733 \times 10^{-2} \\ - & 2.640 \times 10^{-1} & -6.145 \times 10^{-3} \\ - & - & 2.227 \times 10^{-3} \end{pmatrix} \quad (5.6)$$

In the basis of  $D_M/r_{\text{drag}}$ ,  $D_H/r_{\text{drag}}$ , and  $f\sigma_8$  we have:

$$\mathbf{C}_\lambda = \begin{pmatrix} D_M/r_{\text{drag}} & D_H/r_{\text{drag}} & f\sigma_8 \\ 7.709 \times 10^{-1} & -5.656 \times 10^{-2} & 1.750 \times 10^{-2} \\ - & 2.640 \times 10^{-1} & -6.204 \times 10^{-3} \\ - & - & 2.308 \times 10^{-3} \end{pmatrix} \quad (5.7)$$

The results using  $D_M/r_{\text{drag}}$ ,  $Hr_{\text{drag}}$ , and  $f\sigma_8$  basis can be found in the Appendix B.1.

Table 5.1 A summary of the parameter space  $\lambda$ . A flat prior is applied to all parameters with uniform distribution inside the limits and zero otherwise. In all cases, the distortion parameters  $q_\perp$ ,  $q_\parallel$  and  $f\sigma_8$  are all free. We vary also the three bias parameters and  $a_{\text{vir}}$  for the RSD effect at small scale.  $\sigma_{\text{zerr}}$  and  $f_s$  are left optional depending on the feature of the problem we study.

Parameter	Description	Units	Prior limits	
			DR16	DR14
$b_1$	Linear bias	—	[0.25, 6]	[0.25, 6]
$b_2$	Second order bias	—	[-2, 3]	[-1, 6]
$\gamma_3^-$	non local bias	—	[-2, 2]	—
$a_{\text{vir}}$	FoG kurtosis	—	[0.2, 10]	[0.2, 5]
$\sigma_{\text{zerr}}$	Redshift error	Mpc/h	[0, 6]	[0, 6]
$f_s$	fiber collision		[0.2, 0.7]	—
$q_\perp$	Distortion $\perp$ L.O.S	—	[0.5, 1.5]	[0.5, 1.5]
$q_\parallel$	Distortion $\parallel$ L.O.S	—	[0.5, 1.5]	[0.5, 1.5]
$f\sigma_8$	growth parameter	—	[0, 1]	[0, 1]

### 5.1.2 Results in the configuration space: BAO-only analysis

We apply the methodology described in Section 4.5 to the measured eBOSS quasar monopole and quadrupole,  $\xi_{\ell=0,2}(s)$ , in the range  $50 h^{-1}\text{Mpc} < s < 150 h^{-1}\text{Mpc}$ , with bin size  $\Delta s = 5 h^{-1}\text{Mpc}$ . The constraints in the basis of AP parameters can be found in Table 5.3. The

error bars given here are only derived statistically from the MCMC chain with the correction factor  $\sqrt{M} = 1.010$ .

The  $\chi^2$  value of our fit to  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  is  $\chi^2/\text{dof} = 34.1/30$ , while for the fit to  $\alpha_{\text{iso}}$  we have  $\chi^2/\text{dof} = 16.5/15$ . The best fitting models are presented in Fig. 4.8. The top panel shows the monopole, where we fit  $\alpha_{\perp}$  and  $\alpha_{\parallel}$ . However, the model where we fit  $\alpha_{\text{iso}}$  looks almost identical.

We show the likelihood in Fig. 5.4, in terms of  $\Delta\chi^2$ , for our fit to  $\alpha_{\text{iso}}$ . Our BAO measurement is shown by the solid curve, while the dashed curve is the result for a fit to a template that does not include the BAO. This highlights the significance of the BAO feature in the eBOSS DR16 quasar data, as we find that the BAO model is preferred by a significance greater than  $6\sigma$ .

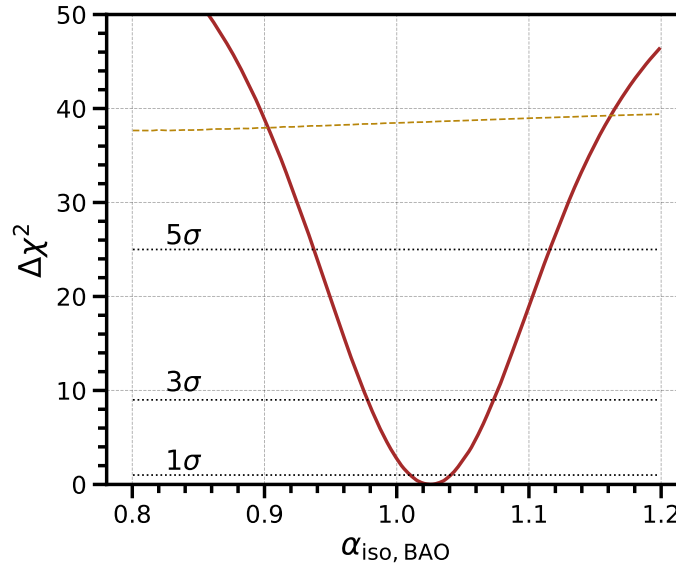


Fig. 5.4 The likelihood of the BAO parameter  $\alpha_{\text{iso}}$  from the fit to the correlation function monopole, in terms of  $\Delta\chi^2$  (solid curve). The dashed curve indicates the likelihood for a model with no BAO feature. The no BAO model has a  $\chi^2$  greater than 37 over the full range of  $\alpha_{\text{iso}}$  values. This implies that the clustering of the eBOSS DR16 quasar sample has a BAO feature at greater than  $6\sigma$  significance.

We convert the BAO  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  results to constraints on the comoving angular diameter and Hubble distance:

$$D_{\text{M}}(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.82(82)(21), \quad (5.8)$$

$$D_{\text{H}}(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 13.22(56)(14). \quad (5.9)$$

The numbers in the first parentheses denote the statistical error, while the second parentheses denotes the error inferred from the mock challenge based on Table 5.2. We do not include the error for observational systematics for the BAO-only fit.

Table 5.2 Systematics for full-shape and BAO-only fit to the mocks. On the OUTERRIM HOD mocks, we took the set with realistic redshift smearing and catastrophic redshift failure for the non-blind mocks. For the blind mocks, we use 24 boxes. The rms is calculated from the EZMOCKS sets with all observational systematics included for the full-shape analysis.

Full-shape	OUTERRIM		EZMOCKS
rms	non-blind	blind	all-syst
$\delta D_M/r_{\text{drag}}$	0.070	0.210	0.104
$\delta D_H/r_{\text{drag}}$	0.057	0.145	0.057
$\delta f\sigma_8$	0.008	0.011	0.004
$\delta f\sigma_{12}$	0.007	0.009	0.004
BAO-only	OUTERRIM		
rms	non-blind	blind	
$\delta D_M/r_{\text{drag}}$	0.133	0.161	
$\delta D_H/r_{\text{drag}}$	0.091	0.113	

Fig. 5.5 compares the posterior distribution of the AP parameters for full-shape and BAO-only analysis. The BAO-only measurements are in good agreement (within  $0.5\sigma$ ) with the full-shape measurements presented in the previous subsection. The degeneracy direction of  $\alpha_\perp - \alpha_\parallel$  for the BAO-only fit (blue contour) is precisely predicted in [Ross et al. \(2015b\)](#). The full-shape measurement is expected to obtain improved results on  $D_M(z)$  and  $D_H(z)$  through the broad-band modeling of the AP effect. For our results, this manifests as a 14 per cent improvement in the statistical uncertainty on  $D_H(z)$ .

It has been shown that the BAO-only analysis is robust to the assumption of fiducial cosmology ([Carter et al., 2019](#)), while the full-shape analysis is potentially sensitive to the shape of the model template. We have performed a detailed analysis using the set of OUTERRIM mocks in blind cosmologies ([Smith et al., 2020](#)) and thus believe the full-shape results, with the inclusion of our systematic uncertainties, are robust to these concerns. The good agreement between the full-shape and BAO-only results further strengthen our confidence. Our BAO results are used, after being combined with those of [Neveux et al. 2020](#), for the cosmological tests in [Müller et al. \(2020\)](#) that only use BAO information.

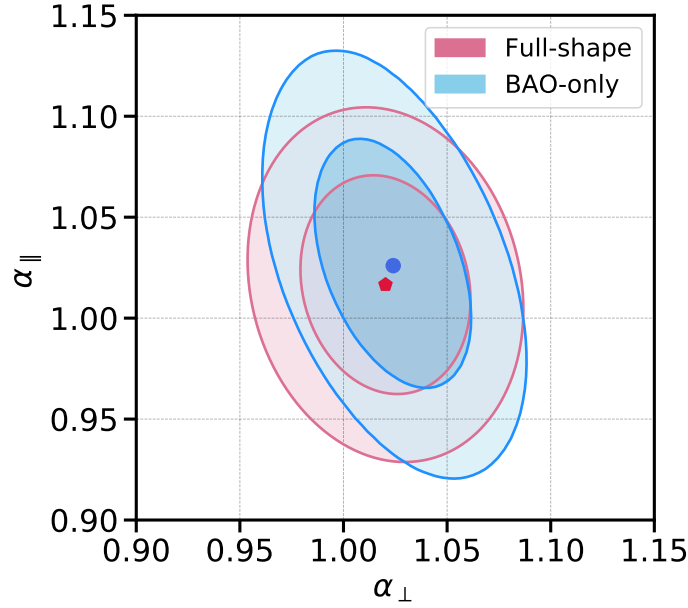


Fig. 5.5 Comparison of full shape (red contour) and BAO only (blue contour) fits in configuration space.

### 5.1.3 Combination of the configuration space and Fourier space results

We use the method described in (Sánchez et al., 2017b) to combine the results. The aim is to compress the information obtained from  $m$  different methods into a single set of measurements, by matching the chi-square calculated from different methods. Under the Gaussian assumption, such measurement should always be possible and we should be able to write down the equation,

$$\mathbf{D}_c = \Psi_c^{-1} \sum_{i=1}^m \left( \sum_{j=1}^m \Psi_{ji} \right) \mathbf{D}_i, \quad (5.10)$$

where the compressed precision matrix is,

$$\Psi_c^{-1} \equiv \left( \sum_{i=1}^m \sum_{j=1}^m \Psi_{ij} \right)^{-1} \quad (5.11)$$

In the case the two methods are completely independent from each other, the big precision matrix,  $\Psi_c$ , reduces to be block diagonal. The statistical error of the data is directly calculated from the MCMC chain. We use the the 1000 EZMOCKS including the systematic effects to estimate the correlation  $C_{m_i(\lambda_n) \otimes m_j(\lambda_l)}$  between the cosmological parameters  $\lambda = \{\lambda_1, \lambda_2 \dots \lambda_n\}$  among different methods  $m = \{m_1, m_2 \dots m_i\}$  as well as the correlation coefficients between cosmological parameters of the same method  $C_{m_i(\lambda_n \otimes \lambda_l)}$ . The estimation of the correlation



between the parameters of the same method is different from the original proposal and we discuss the difference in Section 5.2.

The diagonal elements from the real data are rescaled using the  $\sqrt{M} = 1.036$  factor, which turns out to be identical in both configuration and Fourier space. Since the covariance matrix of the EZMOCKS is estimated from the scattering of the best fit parameters, no correction is needed. Fig. 5.6 shows the correlation coefficients between two methods, with the diagonal terms of the off-diagonal blocks being 0.743, 0.783, 0.844.

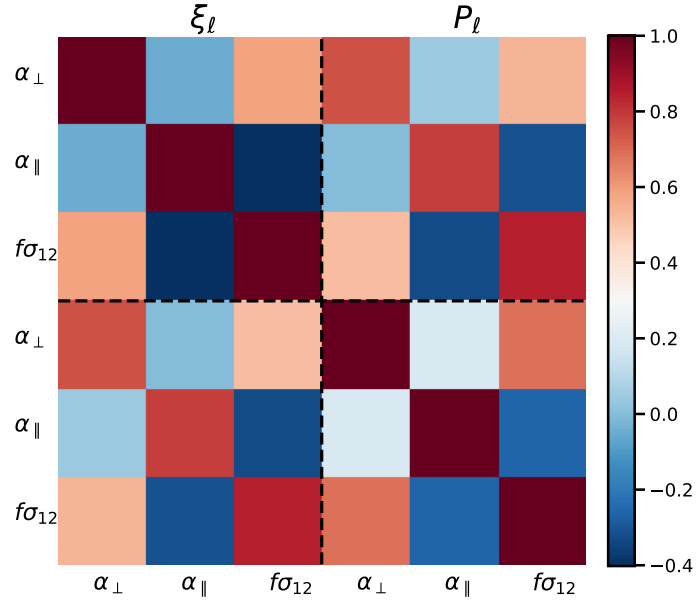


Fig. 5.6 Correlation coefficients for the configuration multipoles and the power spectrum.

Fig. 5.7 shows the posterior for  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$ , and  $f\sigma_{12}$  in configuration space (green), Fourier space (orange), and the combined results using the method described in [Sánchez et al. \(2017b\)](#). The black solid ellipses represent the combined constraints at the 68 and 95 confidence limits. As summarized in Table 5.3, by combining the configuration and Fourier space results, we find an improvement in the statistical uncertainty of  $\sim 7\%$   $\sigma$  in  $\alpha_{\perp}$ ,  $\sim 3\%$   $\sigma$  in  $\alpha_{\parallel}$ ,  $\sim 5\%$   $\sigma$  in  $f\sigma_{12}$ .

To quantify the combined systematic error, we use the non-blind mocks that include the effects of redshift smearing and catastrophic redshifts and the blind mocks with various implementations of HODs (see Section 4.4.2). We firstly combine the configuration and power spectrum multipoles for each of the boxes, and we calculate the correlation coefficients using the 100 realizations for each box. The systematic error is derived from the rms of the difference with respect to the true cosmology. The combined statistics on the OUTERRIM mocks is summarized in Table 5.4. The observational systematics inferred from

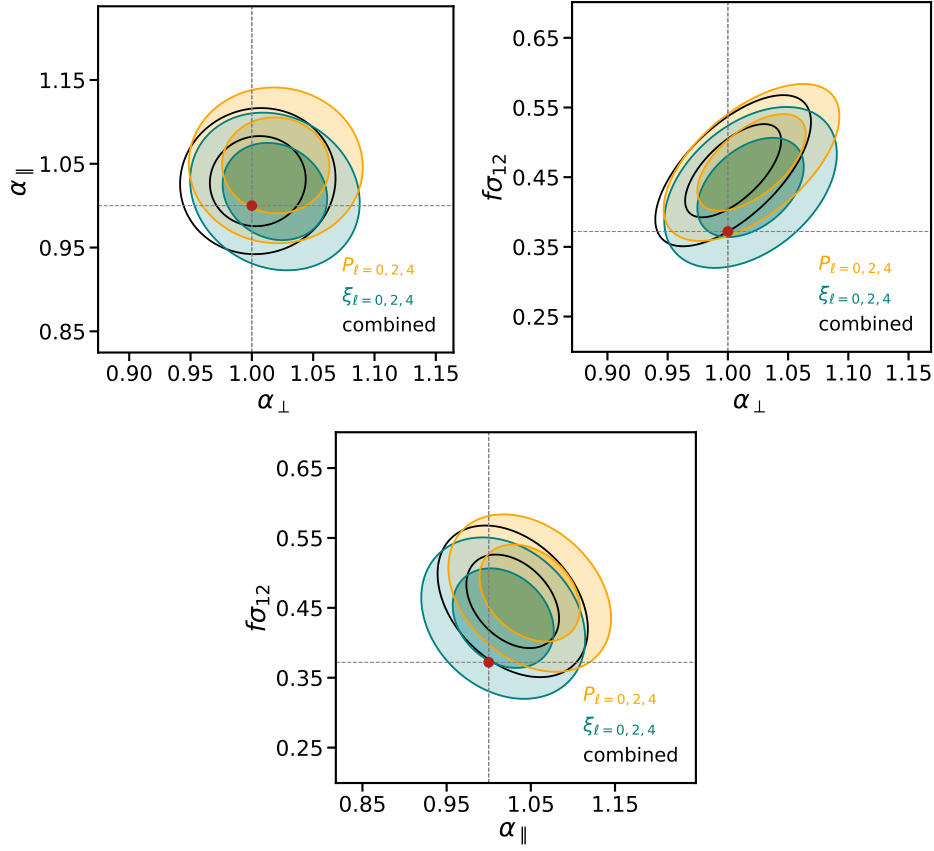


Fig. 5.7 Posterior for  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$  and  $f\sigma_{12}$  configuration space, Fourier space and the combined results using the method described in (Sánchez et al., 2017b). The filled contours are derived from mcmc chains for configuration space (green), and k-space (orange). The black solid ellipses are the combined constraints at 68, 95 confidence limit. The red points denote the values that are inferred from the combined Planck 2018 and BAO results (Collaboration et al., 2018).

Table 5.3 Table summarizes the values on the final DR16 data for the  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$  and  $f\sigma_{12}$  in configuration space, Fourier space and the combined results.

Full-shape	$\alpha_{\perp}$	$\alpha_{\parallel}$	$f\sigma_{12}$
$\xi_{\ell}$	$1.019 \pm 0.028$	$1.017 \pm 0.036$	$0.435 \pm 0.046$
$P_{\ell}$	$1.020 \pm 0.029$	$1.049 \pm 0.038$	$0.471 \pm 0.045$
combined	$1.004 \pm 0.026$	$1.029 \pm 0.036$	$0.460 \pm 0.043$
BAO-only	$\alpha_{\perp}$	$\alpha_{\parallel}$	$\alpha_{\text{iso}}$
$\xi_{\ell}$	$1.024 \pm 0.026$	$1.026 \pm 0.042$	$1.026 \pm 0.016$

the EZMOCKS are directly added to the diagonal terms of the data covariance matrix. Finally, we arrive at the combined result in terms of comoving angular diameter distance, Hubble

parameter, and  $f\sigma_{12}$ ,

$$D_M^c(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.21 \pm 0.79, \quad (5.12)$$

$$D_H^c(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 13.23 \pm 0.47 \quad (5.13)$$

$$f\sigma_{12}^c(z_{\text{eff}} = 1.48) = 0.458 \pm 0.044, \quad (5.14)$$

where the errors include both the statistical and systematic uncertainties. The final covariance matrix for the combined data reads as the following

$$\mathbf{C}_\lambda^c = \begin{pmatrix} D_M/r_{\text{drag}} & D_H/r_{\text{drag}} & f\sigma_{12} \\ 6.227 \times 10^{-1} & 1.424 \times 10^{-2} & 2.235 \times 10^{-2} \\ - & 2.195 \times 10^{-1} & -7.246 \times 10^{-3} \\ - & - & 1.958 \times 10^{-3} \end{pmatrix} \quad (5.15)$$

In addition, in the  $f\sigma_8$  basis we have:

$$D_M^c(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.21 \pm 0.79, \quad (5.16)$$

$$D_H^c(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 13.23 \pm 0.47 \quad (5.17)$$

$$f\sigma_8^c(z_{\text{eff}} = 1.48) = 0.462 \pm 0.045, \quad (5.18)$$

with the covariance matrix:

$$\mathbf{C}_\lambda^c = \begin{pmatrix} D_M/r_{\text{drag}} & D_H/r_{\text{drag}} & f\sigma_8 \\ 6.227 \times 10^{-1} & 1.424 \times 10^{-2} & 2.257 \times 10^{-2} \\ - & 2.195 \times 10^{-1} & -7.315 \times 10^{-3} \\ - & - & 2.020 \times 10^{-3} \end{pmatrix} \quad (5.19)$$

## 5.2 Robustness tests on the data analysis

In this section we describe the various systematic tests we perform on the data to check the robustness of our inferred cosmological constraints. We will consider alternative definitions of the systematics weights, model for the two-point correlation function, definition on the effective redshift, and the impact of the fibre collision correction. The final results are

Table 5.4 Table summarizes the combined systematics for configuration space result and the fourier space results. For the non-blind mocks, we took the set with catastrophic redshift failure. For the blind mocks, we use 24 boxes (without mock6, which includes the catastrophic redshift failure). We use the results from the set with the free  $\sigma_{\text{zerr}}$  parameter and calibrate the offset from the nonblind mocks. For the Ezmocks, we use the set including the full observational systematics, with the fiber collision correction in the model.

rms	non-blind	blind
$\delta D_{\text{M}}/r_{\text{drag}}$	0.079	0.129
$\delta D_{\text{H}}/r_{\text{drag}}$	0.053	0.094
$\delta f \sigma_8$	0.009	0.008
rms	non-blind	blind
$\delta D_{\text{M}}/r_{\text{drag}}$	0.079	0.125
$\delta H r_{\text{drag}}$	0.053	0.094
$\delta f \sigma_{12}$	0.009	0.007

summarised in Table 5.6, which shows how the final measurements of  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$  and  $f\sigma_{12}$  shift with different choices for the systematic corrections.

### 5.2.1 List of tests performed on the data

#### Redshift efficiency weights

The redshift detection efficiency depends, for example, on the signal-to-noise ratio (SNR) of the spectrograph, observational conditions, position of the objects with respect to the focal plane, and the intrinsic properties of the objects. To account for the inhomogeneity in the redshift detection efficiency, we identify the trends in  $n_{\text{good}}/n_{\text{total}}$  as a function of the fibre number ID and the spectral SNR, where  $n_{\text{good}}$  stands for the number of good objects and  $n_{\text{total}}$  is for the total objects. An inverse weighting is assigned to each object to correct for the trend. As discussed in Section 1.3, the efficiency in detecting the redshift of the objects is not uniform across different fibres (see figure 4 in the companion paper [Ross et al. \(2020\)](#)). The detection efficiency is lower near the edge of the CCDs, as well as near the locations of the CCD amplifiers. While the trend as a function of the spectral SNR is weak for the quasar sample, we include the correction to remove any dependency. Both effects are accounted for in the final redshift failure weighting. In [Ata et al. \(2017\)](#) the correction was performed by up-weighting objects by the success rate of the sectors. In Table 5.6 we show the impact of weighting based on the success rate of each sector (denoted as “ $w_{\text{noz,ssr}}$ ”). In addition, we

also show a weighting that only corrects for the trend in fibre ID number, without considering the spectral SNR (denoted as “ $w_{\text{noz,id}}$ ”).

### Photometric weights

In the previous data release 14, we calibrated the trend in  $n_{\text{data}}/n_{\text{random}}$  against the extinction corrected  $g$ -band depth and the extinction coefficients  $E(B - V)$ . In fact, our data also shows trends in the sky background and seeing, in the  $i$ -band (see figure 9 of [Ross et al. \(2020\)](#)). In the final data catalogue we correct for all of these trends. In Table 5.6, we show the impact of using photometric weights that omit the trends in the  $i$ -band, which we denote as “ $w_{\text{photo,no-}i}$ ”.

### Close pair correction

The finite radius of the fibre leads to objects in close pairs being missed, due to fibre collisions. Our fibre collision correction, which models the impact on the two-point correlation function, is described in Section 4.4.3. A more rigorous treatment of this effect can be found in [Bianchi & Percival \(2017\)](#) and [Mohammad et al. \(2018\)](#), where correlation function measurements are corrected using pairwise inverse probability (PIP) weights. The idea is to up-weight the pair counts based on the probability that each pair can be observed. This probability is inferred by running the fibre assignment algorithm many times (on the corresponding eBOSS input target catalogue) to find how often each pair can be observed. The detailed description of catalogue with PIP weights that we use can be found in [Mohammad et al. \(2020\)](#). Table 5.6 shows the impact of using the PIP weighting, which is denoted as  $w_{\text{cp,pip}}$ .

### Impact on the combination of NGC and SGC

We compare two methods for combining the data from the NGC and SGC. In the first method, which is done in our final analysis, the pair counts from the north and south caps are combined. In the second method, the north and south caps are fitted separately, and the posterior distributions are combined. Given that the north and south are statistically independent, the second method would correspond to applying fits simultaneously to both caps, but leaving all the fitting parameters free at the same time (including the AP parameters, growth rate parameter, bias parameters, etc). To determine the correction factor of the fibre collision,  $f_s$  (see Section 4.4.3), for the north and south caps, we measure the projected correlation function. To increase the signal to noise ratio, we integrate over the full depth of the QSO sample along the radial direction. Fig. 5.8 shows the projected correlation function for the NGC, SGC and the combination. When fitting the NGC and SGC separately, we find  $f_s|_{\text{NGC}} = 0.36$  and  $f_s|_{\text{SGC}} = 0.45$ , which is consistent with [Neveux et al. \(2020\)](#). In Table 5.6

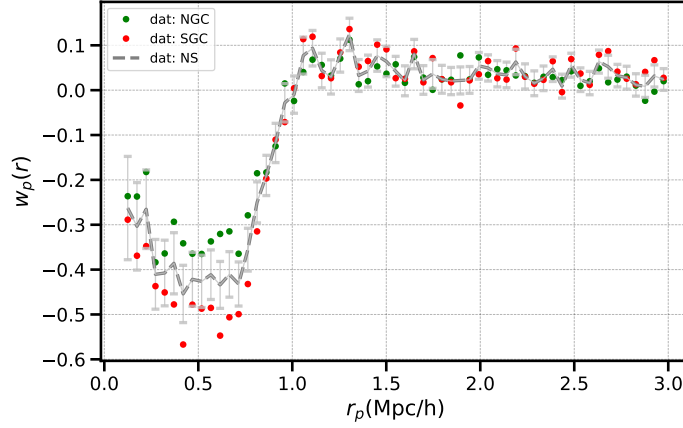


Fig. 5.8 Projected correlation function on the data for north cap (green), south cap (red), and the combined pair counts (grey). The errorbars are derived from the EZmocks.

we show the effect on our results of combining independent fits to the NGC and SGC. The shifts are small compared to the total systematic uncertainty.

### Alternative estimation of the correlation coefficients

As discussed in Section 5.1.3, to estimate the correlation between cosmological parameters measured using different methods, our only option is to use the 1000 EZMOCKS, which include the systematic effects. To estimate the correlation between cosmological parameters within the same method, we have two options: use either the EZMOCKS or use a covariance matrix that is inferred directly from the data. The latter option is justified if, on average, the error inferred from a single realization matches that from the ensemble of the mocks.

Fitting the 1000 EZMOCKS in configuration space, we arrive at the standard deviation (median value) of the 1000 realizations being  $\overline{\text{std}(\alpha_{\perp})} = 0.0398$ ,  $\overline{\text{std}(\alpha_{\parallel})} = 0.0532$  and  $\overline{\text{std}(f\sigma_{12})} = 0.0489$ , which is in good agreement with the scatter in the best fitting parameters for the 1000 realizations ( $\delta\alpha_{\perp} = 0.0405$ ,  $\delta\alpha_{\parallel} = 0.0530$ ,  $\delta f\sigma_{12} = 0.0470$ ). For presenting the results, we select the first option of estimating the correlation coefficients using the mocks. Although the correlation coefficients are cosmology dependent, the estimation from an ensemble of mocks should be more robust and less sensitive to statistical fluctuations. The effect of choosing the second option of using the data to infer the correlation coefficients is shown in Table 7. 5.6.

### Impact of the effective redshift definition

We define the effective redshift,  $z_{\text{eff}}$ , of the quasars using Eq. (5.1), which matches the definition used for the other eBOSS tracers. The main motivation for this definition is that, in

practice, the correlation function is measured in terms of weighted pairs. Taylor expanding the correlation function about  $z_{\text{zeff}}$  gives

$$\xi(z) = \xi(z_{\text{zeff}}) + \left. \frac{d\xi}{dz} \right|_{z=z_{\text{zeff}}} (z_{\text{zeff}} - z) + O[(z_{\text{zeff}} - z)^2], \quad (5.20)$$

and we effectively measure the correlation function

$$\hat{\xi} = \frac{\sum_{i,j} \xi(z) |_{z=(z_i+z_j)/2} w_i w_j}{\sum_{i,j} w_i w_j}, \quad \text{for } \Delta s_{i,j} \in [s_{\min}, s_{\max}]. \quad (5.21)$$

The first order term in Eq. (5.20) vanishes if we define the effective redshift as in Eq. (5.1). However, there is some ambiguity in the definition of the effective redshift. Firstly, objects at higher redshifts are more likely to receive a larger weight, and the effect of this should in principle also be taken into account. Secondly, there is also potential ambiguity in the range of pair separations,  $\Delta s_{i,j}$ , that are summed over. In Table 5.5 we list the value of  $z_{\text{zeff}}$  obtained using different definitions. We find that the range of pair separations has only a marginal impact, while the pair-defined  $z_{\text{zeff}}$  differ by  $\sim 3\%$  compared to the definition we used for the DR14 analysis. Nevertheless, we compare the inferred cosmological parameters obtained using either of the effective redshift definitions, and the difference is small compared to the statistical error (see Table 5.6).

Table 5.5 Effective redshift,  $z_{\text{zeff}}$ , of the NGC, SGC, and combined NGC+SGC, for different definitions of  $z_{\text{zeff}}$ . The first row uses the definition of  $z_{\text{zeff}}$  used in the DR14 analysis. The second and third rows show the definition used in our DR16 analysis, with different ranges of pair separations.

$z_{\text{eff}}$	NGC	SGC	NS
$\frac{\sum_i (w_i * z_i)}{\sum_i w_i}$	1.512	1.520	1.515
$\frac{\sum_{i,j} (w_i w_j * (z_i + z_j)/2)}{\sum_{i,j} w_i w_j}$ $s \in [25, 120] \ h^{-1}\text{Mpc}$	1.474	1.491	1.480
$\frac{\sum_{i,j} (w_i w_j * (z_i + z_j)/2)}{\sum_{i,j} w_i w_j}$ $s \in [20, 160] \ h^{-1}\text{Mpc}$	1.474	1.491	1.480

### 5.2.2 Summary of the robustness test

Table 5.6 shows how the measurements of  $\alpha_{\perp}$ ,  $\alpha_{\parallel}$  and  $f\sigma_{12}$  are shifted, for alternative choices of weighting schemes, compared to the one used in the final data catalogue. In the spectroscopic weighting, the effect of correcting for the trend in the spectral SNR has a

marginal impact on the parameter constraints. In addition, the difference when using the “SSR” weights applied to the DR14 data is at the sub-percent level compared to the statistical error. Similarly, the correction in the photometric weights by including the sky background and seeing in the  $i$ -band also induces changes at a sub-percent level, and therefore this does not influence the conclusions drawn from the DR14 release. The close pair correction using the PIP algorithm has a larger impact on  $\alpha_{\parallel}$  and  $f\sigma_{12}$ , where the latter one accounts for 44% of the statistical error. Given the statistical properties of the two close-pair treatment schemes, this difference is statistically not significant, nevertheless, it would be worth exploring for future denser samples. The table also lists miscellaneous tests including the impact of setting  $f_s = 0$  in our modelling of the fibre collision effect, a different definition of the effective redshift, constraints derived using the gRPT model, a different method to combine the NGC with SGC, and an alternative estimation of the correlation coefficients. These tests all show a much smaller variation compared to the statistical uncertainty, which demonstrates the robustness of our analysis.

Table 5.6 Table compares the impact on potential systematics, that includes alternative weighting schemes for redshift efficiency weights ( $w_{\text{noz},\{\text{id},\text{ssr}\}}$ ), photometric ( $w_{\text{photo},\text{no-}i}$ ), and fibre collision weight ( $w_{\text{cp},\text{pip}}$ ). The table also includes the definition of the effective redshift (5-th row), the correction on the fibre collision effect (6-th row), the difference using the gRPT model (7-th row), the impact on the combination of the caps (8-th row) and alternative estimation of the correlation coefficients when combining the data (9-th row).

	$\Delta\alpha_{\perp}$	$\Delta\alpha_{\parallel}$	$\Delta f\sigma_{12}$
$w_{\text{noz},\text{id}}$	$0.000 \pm 0.026$	$0.001 \pm 0.037$	$0.000 \pm 0.043$
$w_{\text{noz},\text{ssr}}$	$0.002 \pm 0.026$	$-0.002 \pm 0.036$	$0.004 \pm 0.046$
$w_{\text{photo},\text{no-}i}$	$-0.002 \pm 0.027$	$-0.001 \pm 0.036$	$-0.002 \pm 0.042$
$w_{\text{cp},\text{pip}}$	$-0.007 \pm 0.026$	$0.012 \pm 0.034$	$-0.019 \pm 0.043$
fibre collision			
$f_s = 0$	$0.003 \pm 0.027$	$-0.004 \pm 0.036$	$0.007 \pm 0.044$
$z_{\text{eff}} = 1.52$	$0.000 \pm 0.028$	$-0.001 \pm 0.036$	$0.001 \pm 0.044$
model			
gRPT	$0.002 \pm 0.027$	$-0.001 \pm 0.037$	$0.002 \pm 0.044$
NGC+SGC independent	$-0.004 \pm 0.028$	$0.011 \pm 0.037$	$0.010 \pm 0.043$
correlation coeff.	$0.003 \pm 0.026$	$0.007 \pm 0.035$	$-0.002 \pm 0.043$



## Chapter 6

# RSD and the modified gravity models

In spite of the success of the  $\Lambda$ CDM, this standard cosmological model does not provide a satisfying answer to the origin of the cosmic acceleration. Alternative theories of gravity (commonly referred to as modified gravity theories; MG) can provide an explanation with a similar cosmic expansion history to that in  $\Lambda$ CDM, but with different evolution of the growth rate, usually parametrized as  $f(z) \simeq \Omega_m^\gamma(z)$ , with  $\Omega_m$  being the matter density parameter (Linder & Cahn, 2007). A deviation in the index from  $\gamma = 0.55$  would indicate a different theory with distinctive gravitational evolution compared to GR, and therefore has a direct impact on the anisotropic clustering caused by the RSD effect. The constraints obtained from anisotropic clustering measurements can be used to test models of gravity on large-scales. However, these tests require the validation of our analysis techniques in modified gravity scenarios.

A large number of MG theories have been proposed in the literature (see review, Clifton et al., 2012; Joyce et al., 2015; Koyama, 2016). Here we consider two representative families of the MG models: Hu-Sawicki  $f(R)$  gravity model (Hu & Sawicki, 2007) and the normal branch of the Dvali–Gabadadze–Porrati model (nDGP; Dvali et al. 2000), which are also two of the most well studied MG models. Section 6.1 will give a short introduction to the MG models. Section 6.2 will describe how the structure formation is modified in the alternative models. Section 6.3 will describe the mock catalogue used for the study in the paper. Section 6.4 will discuss the fitting template used to obtain the results. Section 6.5 will show the results and discuss the our conclusion from this study and the future perspectives.

## 6.1 Introduction to modified gravity models

The Horndeski class of models represents the most general modification of GR as a scalar-tensor theory that leads to second order equations of motion (Horndeski, 1974). The  $f(R)$

theories, which involve higher powers of the Ricci scalar  $R$ , are among the most popular models within this class. The basic idea of  $f(R)$  theories is to extend the Einstein-Hilbert action to a more general function that depends on the Ricci scalar,

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} (R + f(R)) + S_m(g_{\mu\nu}, \psi_i), \quad (6.1)$$

with  $S_m$  being the action for the total matter field  $\psi_i$  (baryon and dark matter). As is pointed out in Kobayashi et al. (2011) that the Horndeski theory is equivalent to the generalized Galileons (Deffayet et al., 2011), where the Galileons are scalar field that transforms as  $\phi \rightarrow \phi + b_\mu x^\mu + c$  and the generalized Galileon model is an extension of the covariant Galileon (Deffayet et al., 2009) that has second order equation of motion (Kobayashi et al., 2019).

Dvali-Gabadadze-Porrati Gravity (DGP, Dvali et al. (2000)) is a representative of the higher dimensional theories of gravity, it provides alternative explanation that the cosmic acceleration is due to the weaker gravity at horizon scale. In a  $D$ -dimensions of spacetime, the Newtonian potential by a point source drops as  $1/r^{D-3}$ , where the space-time manifold has dimension  $D = 3 + 1$ . From the solar scale experiments the gravity does not appear on higher dimension than  $D = 4$ , however, extra dimensions could be compacted so that they are hidden from the experiment. The DGP model proposes a four-dimensional braneworld embedded in a five dimensional space-time with the action

$$S = \int_{\text{brane}} d^4x \sqrt{-g} \left( \frac{R}{16\pi G} \right) + \int_{\text{bulk}} d^5x \sqrt{-g^{(5)}} \left( \frac{R^{(5)}}{16\pi G^{(5)}} \right), \quad (6.2)$$

where  $g^{(5)}$ ,  $R^{(5)}$  and  $G^{(5)}$  are the equivalents of the four-dimensional quantities in five-dimension bulk. The transition of gravity from 5D to 4D is given by the crossover scale  $r_c \equiv \frac{1}{2} \frac{G^{(5)}}{G}$ . The link between the DGP model and the Horndeski theory is that, the DGP model gives rise to a cubic Galileon interaction  $\sim (\partial\phi)^2 \square\phi$  in its four dimensional effective theory (Kobayashi et al., 2019). In the following sections, we will first summarize the structure formation in both of the models, that leads to different structure growth than GR. The constraints obtained from anisotropic clustering measurements provide a means to test models of gravity on large-scales. However, we need to validate the method for different modified gravity models before we draw a conclusion. In the next, we will discuss how the RSD is affected in both of the model, and describe the settings used to test these models. Finally, we will discuss the results and the implication in constraining the modified gravity models in the future surveys.

## 6.2 Structure formation in modified gravity theory

### 6.2.1 $f(R)$ gravity

The modified Einstein equations for  $f(R)$  model can be obtained by varying the action (see Eq. (6.1)) with respect to the metric  $g_{\mu\nu}$ ,

$$G_{\mu\nu} + f_R R_{\mu\nu} - g_{\mu\nu} \left( \frac{1}{2} f(R) - \square f_R \right) - \nabla_\mu \nabla_\nu f_R = 8\pi G T_{\mu\nu}^m, \quad (6.3)$$

where  $\square = \nabla^\mu \nabla_\mu$  is the d'Alembertian,  $f_R \equiv df(R)/dR$  is the new degree of freedom, known as the scalaron field, the larger amplitude of this field  $|f_R|$  determines a larger deviation from GR. In [Hu & Sawicki \(2007\)](#) an explicit form of  $f(R)$  is proposed as

$$f(R) = -m^2 \frac{c_1}{c_2} \frac{(-R/m^2)^n}{(-R/m^2)^n + 1}, \quad (6.4)$$

where  $n$ ,  $c_1$  and  $c_2$  are new dimensionless parameters in this model and a new mass scale is introduced  $m^2 \equiv H_0^2 \Omega_m$ . By specifying  $c_1/c_2 = 6(\Omega_\Lambda/\Omega_m)$  the  $f(R)$  model can approximately mimic the background expansion of the  $\Lambda$ CDM model and leads to the scalaron field,

$$f_R \approx -n \frac{c_1}{c_2^2} \left( \frac{m^2}{-R} \right)^{n+1}. \quad (6.5)$$

It is more common to express the free parameter  $c_1/c_2^2$  in terms of  $f_{R0}$  as the scalaron field of today,

$$\frac{c_1}{c_2^2} = -\frac{1}{n} \left[ 3 \left( 1 + 4 \frac{\Omega_\Lambda}{\Omega_m} \right) \right]^{n+1} f_{R0}. \quad (6.6)$$

Here we focus on the case of  $n = 1$ ,  $f_{R0} = -10^{-6}$  and  $f_{R0} = -10^{-5}$ , referred as F6 and F5. Under the quasi-static (drop the time derivative of the scalaron field) and weak field approximations, the non-linear structure formation in the longitudinal gauge is determined by the modified Poisson equation ([Bose et al., 2015](#)),

$$\nabla^2 \Phi = \frac{16\pi G}{3} a^2 (\rho_m - \bar{\rho}_m) + \frac{1}{6} a^2 (R(f_R) - \bar{R}), \quad (6.7)$$

and the scalaron equation of motion

$$\nabla^2 f_R = -\frac{a^2}{3} [R(f_R) - \bar{R} + 8\pi G (\rho_m - \bar{\rho}_m)] \simeq \frac{dV_{\text{eff}}}{df_R}, \quad (6.8)$$

where the quantity with bar  $\bar{\cdot}$  denotes the background and we use the approximation  $|f(R)| \ll |R|$  and  $|f_R| \ll 1$ . The effective potential  $V_{\text{eff}}$  can be splitted into a term that purely depends on the scalaron field  $f_R$  and a term that also depends on the density. A common choice is  $V_{\text{eff}} = f_R^{-n} + e^{\alpha f_R/M_{\text{Pl}}}$ , with  $\alpha$  being the coupling strength between the scalaron field and the matter.

### 6.2.2 Dvali–Gabadadze–Porrati model

The background expansion rate of the DGP model can be derived from the variation of the action Eq. (6.2),

$$\frac{H(a)}{H_0} = \sqrt{\Omega_m a^{a-3} + \Omega_\Lambda(a) + \Omega_{\text{rc}} \pm \sqrt{\Omega_{\text{rc}}}}, \quad (6.9)$$

there exist two branches in the DGP model, the  $(-)$  sign denotes the self-accelerating branch and the  $(+)$  sign denotes the normal branch. In the self-accelerating branch the Universe expansion accelerates without a dark energy component  $\Omega_{\text{de}} = 0$ . However, this branch suffers from theoretical instabilities, with negative sign for the mass term (known as a ghost), and has been ruled out by CMB and supernovae data (Fang et al., 2008). Therefore, we focus only the normal branch (nDGP), which requires a non vanishing dark energy  $\Omega_{\text{de}}(a) \neq 0$  to derive an accelerated expansion and thus being theoretically less appealing. Nevertheless, it satisfies a series of constraints tests and is a very useful toy model to understand the deviation of GR using the large scale structure.

The structure formation in the nDGP model is governed by the following equations (Koyama & Silva, 2007),

$$\nabla^2 \Phi = 4\pi G a^2 \delta \rho_m + \frac{1}{2} \nabla^2 \varphi, \quad (6.10)$$

$$\nabla^2 \varphi + \frac{r_c^2}{3\beta_{\text{DGP}} a^2} \left[ (\nabla^2 \varphi)^2 - (\nabla_i \nabla_j \varphi)^2 \right] = \frac{8\pi G a^2}{3\beta_{\text{DGP}}} \delta \rho_m, \quad (6.11)$$

where  $\psi = \Phi + \Psi$  is a scalar degree of freedom referred to the brane bending modes, the perturbation of the non-relativistic matter is defined as  $\delta \rho_m = \bar{\rho}_m - \rho_m$ .  $\beta_{\text{DGP}}(a) = 1 + 2Hr_c \left(1 + \frac{\dot{H}}{3H^2}\right)$ .

Finally, the linear growth for the matter fluctuations in these gravity models can be obtained by solving the equation of the linear growth factor,  $D(k, \tau)$ , and we have

$$D'' + \left(2 - \frac{3}{2} \Omega_m(a)\right) D' - \frac{3}{2} \frac{G_{\text{eff}}}{G} \Omega_m(a) D = 0, \quad (6.12)$$

where  $\prime$  denotes a derivative with respect to  $\ln a$  and  $G_{\text{eff}}$  takes the values,

$$\frac{G_{\text{eff}}}{G} = \begin{cases} 1 & \text{GR} \\ 1 + k^2 / [3(k^2 + a^2 m_{f_R}^2)] & f(R) \\ 1 + 1 / [3\beta_{\text{DGP}}(a)] & \text{nDGP} \end{cases} \quad (6.13)$$

where  $m_{f_R} \simeq [3f_{RR}]^{-2}$  is the mass of the scalaron field, for  $f(R)$  the linear growth of structure is scale dependent in contrast to DGP and GR.

### 6.2.3 Screening mechanism

Modifications to general relativity typically need to include screening mechanism, so that the additional degrees of freedom in dense regions is suppressed in order to recover the solar system constraints.

The  $f(R)$  theory employs the chameleon mechanism (Khoury & Weltman, 2004) to suppress the enhancement of the fifth force, where the force is proportional to the gradient of the scalar field  $F_{5\text{th}} \propto \vec{\nabla} f_R$ , and the dynamics of the field is given by Eq. (6.8). The solution to the equation of motion of the scalaron field shows that the field is suppressed by the mass of the field as  $\propto e^{-m_{f_R} r}$ , as the mass is defined as

$$m_{f_R}^2 \equiv \frac{d}{df_R} \left( \frac{dV_{\text{eff}}}{df_R} \right) \propto n(n+1)(-R)^{n+2}. \quad (6.14)$$

The minimum effective potential occurs when  $R = -8\pi G\rho_m$  by minimizing the second equality of Eq. (6.8), and it is easy to see that deep in the overdense regime the mass becomes larger and therefore the scalaron field is more suppressed.

The DGP theory employs Vainshtein screening mechanism (Vainshtein, 1972). In this case, the gradient of the field can be obtained by integrating Eq. (6.11), and the corresponding solution is

$$\varphi_{,r} = \frac{4}{3\beta_{\text{DGP}}} \left( \frac{r}{r_V} \right)^3 \left[ -1 + \sqrt{1 + \left( \frac{r_V}{r} \right)^3} \right] \frac{GM(r)}{r^2}, \quad (6.15)$$

where the Vainshtein radius is given by

$$r_V(r) = \left( \frac{16r_c^2 GM(r)}{9\beta_{\text{DGP}}^2} \right)^{1/3}, \quad (6.16)$$

it's clear to see that for  $r_V \gg r$  the effects of fifth forces are suppressed.

### 6.3 Construction of the mock catalogues

In [Hernández-Aguayo et al. \(2019\)](#) we wanted to answer the following questions: firstly, what accuracy can a current leading cosmological BOSS-like survey provide to constrain the MG models using the RSD. Furthermore, is it sufficient to use a standard  $\Lambda$ CDM template of the two point correlation function to fit for the MG models given the reasonable estimates of theoretical and observational uncertainties? To answer these questions, the dark matter only N-body simulations (ELEPHANT ; Extended LEnsing PHysics using ANalytic ray Tracing) performed on the ECOSMOG code ([Li et al., 2012](#)) for  $f(R)$  and ([Li et al., 2013](#)) for DGP are used. Both of the codes are modified versions of the publicly available N-body and hydrodynamical simulation code RAMSES ([Teyssier, 2002](#)) and they inherit the efficient MPI parallelisation and the adaptive-mesh-refinement (AMR). The cosmological parameters are the best-fit values from the WMAP9 collaboration ([Hinshaw et al., 2013](#)),

$$\{\Omega_b, \Omega_{\text{CDM}}, h, n_s, \sigma_8\} = \{0.046, 0.235, 0.697, 0.971, 0.82\}. \quad (6.17)$$

The simulation starts at initial redshift  $z_{\text{ini}} = 49$  and evolves until today  $z_{\text{ini}} = 0$ , with the initial conditions generated using the publicly available MPGRAFIC code ([Prunet et al., 2008](#)). The cubic comoving box size is  $L_{\text{box}} = 1024h^{-1}\text{Mpc}$ , with particle number  $N_p = 1024^3$  and mass resolution  $m_p = 7.798 \times 10^{10} h^{-1} M_\odot$ .

The halos were identified using the halo finder ROCKSTAR ([Behroozi et al., 2013](#)) at  $z = 0, 0.3$  and  $0.5$ . The halo mass  $M_{200c}$  is defined as the enclosed sphere of radius  $r_{200c}$ , with density 200 times the critical density of the Universe.

The dark matter halos of the simulations were populated with HOD, as introduced in Sec. 4.4.1 but follows a different parametrization for the mean number of central  $\langle N_c(M) \rangle$  galaxy and satellite galaxy  $\langle N_s(M) \rangle$  ([Zheng et al., 2005, 2007](#)),

$$\langle N_c(M) \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log_{10} M - \log_{10} M_{\text{min}}}{\sigma \log M} \right) \right], \quad (6.18)$$

$$\langle N_s(M) \rangle = \langle N_c(M) \rangle \left( \frac{M - M_0}{M_1} \right)^\alpha. \quad (6.19)$$

The mock catalogues were constructed in real space. The central galaxies are placed at the center of mass of the host halos and take the velocity information of the host halos. The satellite galaxies follow a Navarro-Frenk-White (NFW) profile ([Navarro et al., 1996, 1997](#)) with a uniform angular distribution, the position of satellite galaxies is randomly chosen within the halo radius ( $0 < r < r_{200c}$ ), and their velocity is the halo velocity plus a perturbation along the  $x$ ,  $y$  and  $z$  coordinates drawn from a Gaussian distribution with variance equal to

the 1D velocity dispersion of the host halo. The HOD parameters in Eq. (6.18) and (6.19) are tuned to match the number density and the clustering of the real-space two point correlation function. In the GR case, the values of the parameters are taken from the BOSS CMASS DR9 sample (Manera et al., 2012a):  $\log_{10}(M_{\min}/[h^{-1}M_{\odot}]) = 13.09$ ,  $\log_{10}(M_1/[h^{-1}M_{\odot}]) = 14.00$ ,  $\log_{10}(M_0/[h^{-1}M_{\odot}]) = 13.077$ ,  $\sigma_{\log M} = 0.596$  and  $\alpha = 1.0127$ . The redshift distribution of the CMASS sample  $0.4z < 0.7$  is slightly different from the ones studied in the paper, however, since the goal was to understand the growth rate using galaxy catalogues with similar clustering. The linear bias of the mock is measured as

$$b(r, z) \equiv \frac{\xi_{\text{gg}}(r, z)}{\xi_{\text{gm}}(r, z)}, \quad (6.20)$$

which is less expensive to compute than  $\xi_{\text{gm}}/\xi_{\text{mm}}$ . At sufficiently large scales we expect  $b(r) \approx \text{const.}$ , hence to measure the linear galaxy bias from our mock catalogues we make a fit of Eq. (6.20) to a constant function using data in the range  $r_{\min} \leq r \leq r_{\max}$ , with  $r_{\max} = 150 h^{-1}\text{Mpc}$  and  $10 < r_{\min}/(h^{-1}\text{Mpc}) < 45$ , then we take the mean over all best-fitting values.

## 6.4 Fitting template for two point correlation

To test the standard analysis pipeline, we used the same fitting template as the one used for the BOSS analysis. We explore the likelihood function with the Gaussian distribution. To model the gravitational nonlinear evolution, we used the gRPT model. The redshift space power spectrum and the bias expansion both follow the description in Section 3.2 and Section 3.3.1. As the mocks are designed to match the LRG sample, there is no need to include the Gaussian damping term for modelling the redshift error. We tried both fixing and varying  $\gamma_3^-$  in our fitting. When fixing the parameter using the local Lagrangian relation,  $\gamma_3^- = -\frac{11}{42}(b_1 - 1)$ , we found the linear bias is biased low compared to the true value. There are two possible reasons for this behaviour. Firstly, the linear bias,  $b_1$ , is scale dependent with a contribution  $\nabla\delta$ , which is ignored in our simplified treatment and making the linear bias degenerate with  $\gamma_3^-$ . Secondly,  $\gamma_3^-$  is formulated in the Eulerian coordinate, while the local biasing schemes are compatible with the Lagrangian bias only when matter evolution and structure formation is well within the linear and local regime (Matsubara, 2011). Therefore varying  $\gamma_3^-$  should result in a more accurate value and this conclusion is consistent with the previous findings (Grieb et al., 2017). In the results below we shall always vary  $\gamma_3^-$ . We use the analytical covariance matrix with the Gaussian approximation (Grieb et al., 2016). The input power spectrum is calculated by the nonlinear model based on the best fitting

values obtained from the MCMC chain. Such Gaussian recipe has been tested recently in both Fourier and configuration space by comparing to covariance matrices generated by hundreds of N-body simulations as well as thousands of different fast mock simulations and found them to be in good agreement (Blot et al., 2018; Lippich et al., 2019). At the same time, there are also studies on including the corrections from higher-order statistics and super-sampling mode (Barreira et al., 2018). However, for the scales of interest in this study, there is no sensitivity to these corrections and the Gaussian covariance matrix should be a good approximation. The parameters that enter the default fitting are  $\{f\sigma_{12}, b_1, b_2, \gamma_3^-, a_{\text{vir}}\}$ . When applying the AP test, two additional parameters enter the fitting  $\{\alpha_{\parallel}, \alpha_{\perp}\}$ . Finally, we marginalise over the nuisance parameters to find the probability distribution of the distortion parameter  $\beta = f/b_1$ .

## 6.5 Results and discussion

Fig. 6.2 shows the constraints on  $\beta$  using the nonlinear gPRT+RSD model by running MCMC. The upper panels present the results for the three multipoles ( $\xi_{\ell}(s)$ ,  $l = 0, 2, 4$ ) of the correlation function for two ranges of scales:  $s = 20 - 150 h^{-1}\text{Mpc}$  (left) and  $s = 40 - 150 h^{-1}\text{Mpc}$  (right). For comparison, we display the results of using three wedges ( $\xi_w(s)$ ) in the bottom panels of Fig. 6.2. We observe an overestimation of  $\beta$  for all models at all redshifts when the fit is done using  $s_{\text{min}} = 20 h^{-1}\text{Mpc}$ , for both multipoles and wedges. We have checked the linear bias fitted from the nonlinear model and found it to be in good agreement with the values measured from the mock galaxy catalogues using Eq. (6.20). This suggests that the higher estimation of  $\beta$  comes from the  $f\sigma_{12}$ ; the same conclusion is in agreement with the one found by Barreira et al. (2016). In our case we have rescaled the covariance matrix by the number of simulations and the error bar is therefore smaller than the error bar presented in Barreira et al. (2016). When we used the range scale with  $s_{\text{min}} = 40 h^{-1}\text{Mpc}$ , the constraints are in good agreement with the fiducial values.

We note a slight difference between the results obtained from the multipoles-based estimator and the wedges-based estimator. This is an indication for the non-negligible higher order statistics in the two-dimensional correlation function  $\xi(s, \mu)$ . To further explore this point, we have compared the difference between the multipoles directly measured from the mock catalogues and the multipoles obtained by transforming the measured wedges using Eq. (4.10), and found a difference in the hexadecapole at scale  $\approx 20 h^{-1}\text{Mpc}$  (we do not show the plot here). The difference can lead to different constraints on the nuisance parameters and have consequential impact on the parameter of interest, which is a source for the biased constraints by setting  $s_{\text{min}} = 20 h^{-1}\text{Mpc}$ . For the case of  $s_{\text{min}} = 40 h^{-1}\text{Mpc}$ , the minimum



length scale used in the fitting is larger enough, and the impact of higher-order statistics is smaller, which explains why the agreement is improved at all three redshifts.

In the lower subpanels of Fig. 6.2 we plot the relative differences between the modified gravity models and GR. Similar to the findings using linear theory, we find that the two variants of  $f(R)$  gravity studied here are indistinguishable from GR given the size of the statistical error. While N5 shows stronger deviation from GR, it is also not clearly distinguishable from the latter. N1 is the only one of our four MG models that could be distinct from GR at  $1\text{-}\sigma$  given the statistical uncertainties, and not including systematic errors.

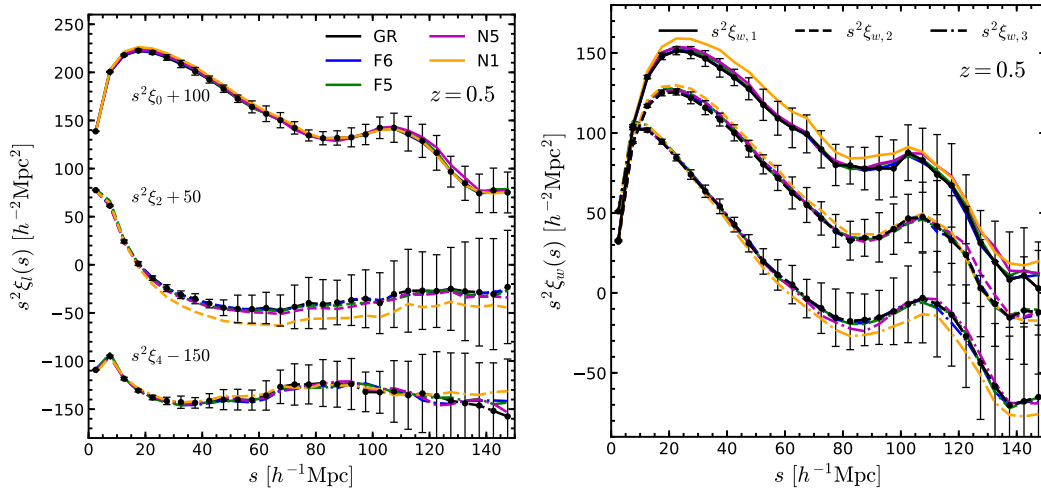


Fig. 6.1 *Left panel*: Monopole, quadrupole and hexadecapole moments of the correlation function for our five gravity models at  $z = 0.5$ . The moments have been shifted by a factor of 100, 50 and  $-150$  for better visualization. *Right panel*: Clustering wedges measured at  $z = 0.5$  for all gravity models as labelled, the upper wedge (solid lines) correspond to angles with  $0 < \mu < 1/3$ , the middle wedge (dashed lines) to  $1/3 < \mu < 2/3$  and the lower wedge (dot-dashed lines) to  $2/3 < \mu < 1$ . The error bars correspond to the standard deviation over fifteen GR measurements.

The results can have a few implications:

First, RSD on linear and mildly-nonlinear scales does not seem to be a great probe of modified gravity, in particular for  $f(R)$  gravity. This conclusion is expected to hold true for other MG models depending on chameleon screening to recover GR in high-density regions, for which the effect of the fifth force is generally restricted to at most  $O(10)$  Mpc (Brax et al., 2012a,b). This conclusion, however, may not apply to RSD on small and highly nonlinear scales, where the velocity field could be significantly enhanced by the fifth force in chameleon models (see, e.g., He et al., 2018). We suspect similar conclusions should hold for the symmetron (Hinterbichler & Khoury, 2010) and dilaton (Brax et al., 2010) models, for which the fifth force is also of the Yukawa type, with an inverse Compton mass of  $\lesssim O(10)$

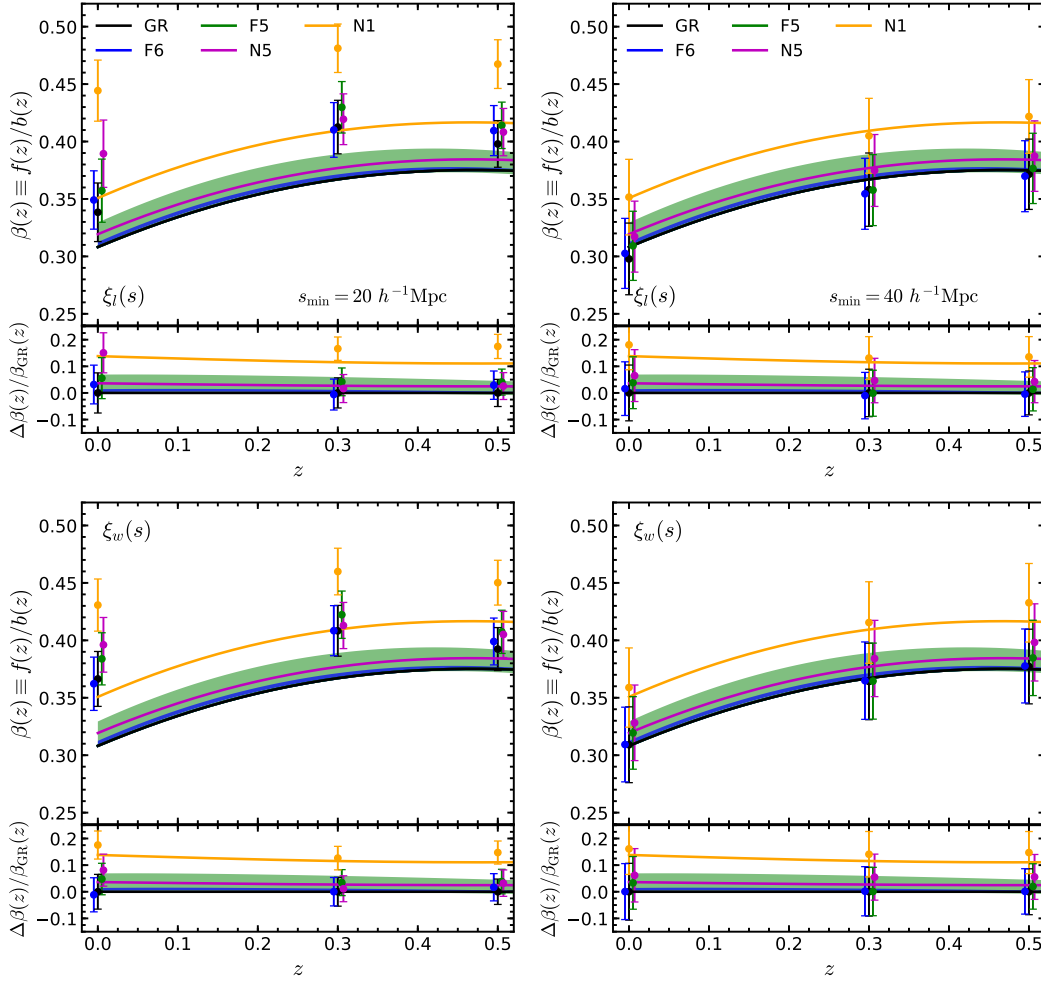


Fig. 6.2 Fits using the nonlinear model. The upper panel shows the multipole moments of the correlation functions,  $\xi_\ell(s)$ . The lower panel shows the clustering wedges  $\xi_{w_i}(s)$ . In both of the plots the AP parameters are treated as free parameters.

Mpc. Vainshtein screening models, such as nDGP, on the other hand, has a fifth force that is non-negligible on large scales, which is why the constraint is stronger.

Second, given the weak constraining power from large-scale RSD and the relatively large scale ( $s_{\min} = 40 h^{-1}\text{Mpc}$ ) needed to get unbiased constraints even for GR, a theoretical model based on linear theory prediction or higher-order perturbation calculation developed for GR does not seem to lead to noticeably biased constraints on the  $\beta$  parameter. This suggests a faster way to explore the MG model and parameter space, at least at the initial stage of delineating models and parameters.

Third, we have explicitly checked that the real-space galaxy correlation functions of the MG models deviate more significantly from GR prediction if the mock galaxy catalogues

were constructed using the same HOD parameters as GR, or if haloes were used instead of HOD galaxies. As argued above, in this study the HOD parameters for MG models are tuned so that the real-space  $\xi_{\text{gg}}$  match between the different models, which is motivated by the fact that there is only one Universe from which the observed  $\xi_{\text{gg}}$  are to be derived, and whatever the cosmological model, it should be required to reproduce such an observation to start with. A more detailed theoretical model of RSD on linear and mildly-nonlinear scales should take this into account. In practice, there is no real-space  $\xi_{\text{gg}}$  from observations to match with, and the HOD parameters are often tuned to match the observed projected two-point correlation function  $w(\theta)$  to get rid of RSD effects: doing this will leave more freedom for the choices of HOD parameters, and we expect it to also lead to larger uncertainties in the constraints on  $\beta$ .

Finally, there are few points could be treated in a more rigorous way. Regarding the modelling of the nonlinear gravitational evolution in the presence of the a fifth-force, [Bose & Koyama \(2016\)](#); [Bose et al. \(2018\)](#); [Taruya \(2016\)](#) incorporated the calculation for both  $f(R)$  and DGP model in the perturbation models. In terms of the large scale bias, as we can see from Eq. (6.12) that, the linear growth factor is not guaranteed to be separable in space and time. In the case of  $f(R)$  theory the linear bias can become a scale dependent quantity,  $b(\tau, \mathbf{k}) = 1 + (b_1(\tau_0) - 1) \frac{D(\tau_0, \mathbf{k})}{D(\tau, \mathbf{k})}$ , with  $b_1(\tau_0)$  being the initial bias at some early time  $\tau_0$  ([Hui & Parfrey, 2008](#)). As discussed in ([Desjacques et al., 2018](#)) on scales larger than Compton wavelength of the scalar field one can perform a perturbative expansion  $k^2/(am)^2$ , and the scale dependence can be absorbed into a higher-derivative bias which is absent in this work. Additionally, the screening mechanism is also a nonlinear process as the Chameleon mechanism employed in the  $f(R)$  model invokes a  $\phi$  dependence in the bias expansion (the scalar field  $\phi(\mathbf{k}) \sim k^2/(am)^2 \Phi(\mathbf{k})$ ). This effect can also be modelled by a higher-derivative bias. Although none of these points in improving the modelling of two-point statistics will change our conclusion in this work, as the impact of  $f(R)$  model on the RSD is very small on large scales. We need to be aware that there exist some well motivated physics are not included.



# Chapter 7

## Summary and outlook

### 7.1 Clustering analysis of the eBOSS QSO sample

For over a decade, cosmology programs within the SDSS have delivered accurate measurements of the expansion and growth of structure histories of the Universe. The completion of the 5-year eBOSS program marks a milestone as part of those efforts. The eBOSS program showed for the first time the detection of BAO using quasars at a mean redshift of  $z_{\text{eff}} \approx 1.5$ . These sparse, bright objects are demonstrated to be robust tracers of the underlying matter field.

During the DR14 analysis, we made use of a state-of-the-art model of non-linear evolution, bias, and RSD (gRPT+RSD model). We projected the information of the full two-dimensional correlation function  $\xi(s, \mu)$  of the eBOSS quasar sample into Legendre multipoles  $\xi_\ell(s, \mu)$  with  $\ell = 0, 2$  or  $\ell = 0, 2, 4$  and clustering wedges measured using two or three  $\mu$ -bins,  $\xi_{2w}(s)$  and  $\xi_{3w}(s)$ . We introduced an extra damping term in the model to account for the quasar redshift uncertainty. Our tests demonstrated that the analysis of the first three nonzero Legendre multipoles provides tighter constraints than the other statistics we considered.

In the analysis of the final eBOSS DR16 quasar sample, we used the Legendre multipoles  $\xi_\ell(s, \mu)$  with  $\ell = 0, 2, 4$ , which yielded the best constraints based on the experience of DR14 analysis. We implemented a new recipe to describe the correlation function, which uses RESPRESSO (Nishimichi et al., 2017) to calculate the matter power spectrum, and use the fitting formulae of Bel et al. (2019) for the power spectra that involve velocities. Our FORTRAN version of the RESPRESSO package will be released in the future, as part of the COMPASS cosmological parameter analysis toolkit.

In the final data release, we double the number of objects and the survey area compared to the DR14 sample, we have a  $6\sigma$  detection of the BAO signal in configuration space, which is consistent with the Fourier space analysis (Neveux et al., 2020). Compared to the DR14

analysis, statistically we gain  $\sim 45\%$  in distance measurement and  $\sim 30\%$  in the growth measurement. We arrive at the comoving angular diameter distance  $D_M(z_{\text{eff}})/r_{\text{drag}} = 30.66 \pm 0.88$ , the Hubble distance  $D_H(z_{\text{eff}})/r_{\text{drag}} = 13.11 \pm 0.52$ , and the cosmic structure growth rate  $f\sigma_{12}(z_{\text{eff}}) = 0.435 \pm 0.048$ . Our analysis in the configuration space combined with the analysis in the Fourier space (Neveux et al., 2020) allow us to obtain a tighter constraints in the cosmological distance and growth rate parameters:  $D_M^c(z_{\text{eff}})/r_{\text{drag}} = 30.25 \pm 0.79$ ,  $D_H^c(z_{\text{eff}})/r_{\text{drag}} = 13.26 \pm 0.47$ , and  $f\sigma_{12}^c(z_{\text{eff}}) = 0.459 \pm 0.045$ .

The measurements of the Alcock Paczynski parameters are found to be within  $1\sigma$  to the  $\Lambda$ CDM model to the best-fitting parameter from Planck combined with BAO (Collaboration et al., 2018). The growth rate measurement  $f\sigma_{12}$  in the configuration space is found to be  $1.4\sigma$  than the  $\Lambda$ CDM model with CMB+BAO constraints. Meanwhile, when combining with the results in the Fourier space, the growth rate  $f\sigma_{12}$  measurement is  $\sim 2\sigma$  higher than the  $\Lambda$ CDM model with CMB+BAO constraints.

We performed extensive tests to quantify potential systematics and focused on testing systematics in the observational effects as well as the modelling of the two-point correlation function. We tested the observational systematics using fast mocks including various angular systematics (such as fibre collision, photometric, and redshift failure effects). We correct for the largest angular systematics (fibre collisions) using a modified version of the recipe of Hahn et al. (2017). We also corrected for the radial integral constraints as described in de Mattia & Ruhlmann-Kleider (2019). Based on these tests, the residual observational systematics on the inferred parameters are shown to be at sub-percent level. Based on a set of HOD mocks built on N-body simulation (Smith et al., 2020), we examine the modelling of the two-point correlation function. In these mocks we checked the impact of various HODs and also include different redshift uncertainty distribution, as well as the catastrophic redshift failure objects (potentially important for the future surveys). The model can very well handle these effects, and we can recover 1 per cent accuracy for the distance measurement and 3 per cent for the growth rate measurement.

A larger systematics turns out to be the impact of the fiducial cosmology and is the dominant source of our systematic error budget that accounts for up to 30 per cent of the statistical error. This is one of the most valuable lessons we learnt from the eBOSS analysis. As a consistency check, we also perform a BAO-only analysis, which was proven to be more robust to the assumption of the fiducial cosmology (Carter et al., 2019). We find good agreement between the full-shape and BAO-only analysis, which demonstrates the robustness of the methodology given the current statistical precision.

In line with our findings from DR14 (Gil-Marín et al., 2018; Hou et al., 2018; Zarrouk et al., 2018), we demonstrated that quasars are robust tracers of the underlying matter field.

The results from this work will enter [Müller et al. \(2020\)](#) to further constrain the cosmological models in combination with other tracers.

## 7.2 RSD in modified gravity theories

Alternative models of gravity modify the dynamics of the gravitational evolution and therefore have a direct impact on the anisotropic clustering caused by the redshift distortion effect. The constraints obtained from anisotropic clustering measurements can be used to test models of gravity on large-scales. However, these tests require the validation of our analysis techniques in modified gravity scenarios.

To do so, in [Hernández-Aguayo et al. \(2019\)](#) we used cosmological dark-matter-only N-body simulations and populated dark matter haloes with galaxies following a HOD prescription. We did this analysis for three low redshift bins, respectively at  $z = 0, 0.3$  and  $0.5$ , because the modified gravity models studied here are expected to deviate from GR more significantly at late times. Since the nature of gravity is different in every model, we tuned the HOD parameters such that essentially every catalogue matches the number density and the real space correlation function measured for the BOSS CMASS DR9 ([Anderson et al., 2012](#); [Manera et al., 2012b](#)). We used the distant-observer approximation to map galaxies from real- to redshift-space coordinates along three line-of-sight directions (chosen to be parallel to the three axes of the simulation box) for each realisation of mock galaxy catalogue. We applied the gRPT+RSD model and used the same pipeline as we analysed the BOSS CMASS sample. We used different estimators to extract information about the RSD distortion parameter  $\beta$  in each model. We have used both the multipole moments  $\xi_\ell(s)$  and the clustering wedges of the correlation function. For both RSD models we performed fits over two ranges of scales,  $s = 20 - 150 h^{-1}\text{Mpc}$  and  $s = 40 - 150 h^{-1}\text{Mpc}$ .

We found that the nonlinear model used here overpredicts the value of  $\beta$  when considering  $s_{\min} = 20 h^{-1}\text{Mpc}$ ; this trend was also found by [Barreira et al. \(2016\)](#) using the same range scales. In the large-scale regime ( $s_{\min} = 40 h^{-1}\text{Mpc}$ ), the nonlinear model is able to recover the true value of  $\beta$ , especially for clustering wedges.

Another study reported by [White et al. \(2015\)](#) found unbiased estimations of the growth rate for scales  $s > 30 h^{-1}\text{Mpc}$ . However, there are some differences between the work presented by [White et al. \(2015\)](#) and ours. First, they fitted only the monopole and quadrupole of the correlation function. Second, opposite to our findings they underpredicted the linear growth rate when considered scales with  $s_{\min} = 20 h^{-1}\text{Mpc}$ .

Our main conclusions are as follows: 1). Measurements of RSD on large scales can help us to distinguish between some gravity models, such as N1, but in general the model

differences from GR are small compared with statistical and theoretical uncertainties, in particular for  $f(R)$  or chameleon models in general. 2). Chameleon and Vainshtein models have distinct predictions, which are directly related to the different properties of the fifth forces incorporated in the model: in chameleon-type models, the fifth force is of Yukawa type and gets suppressed on scales above the inverse Compton mass of the scalar field (typically  $\sim 10$  Mpc or smaller), while for Vainshtein-type models the fifth force is long ranged and can alter the large-scale velocity field substantially. 3). We have tested the effect of using different ranges of scales in the fitting, and found that for  $s_{\min} = 20 h^{-1} \text{Mpc}$  the nonlinear model cannot recover  $\beta$  correctly at all redshifts for all models including GR. In spite of the theory predicting a higher  $\beta$  at the scale range with  $s_{\min} = 20 h^{-1} \text{Mpc}$ , the relative difference between the MG models and GR agrees well with the fiducial values from the simulation. The higher biased  $\beta$  value can be resolved by excluding data between 20 and  $40 h^{-1} \text{Mpc}$ , at the expense of increased uncertainties. 4). Using different estimators such as multipoles and wedges can produce different constraints because of the different information they encode, but the difference is not statistically significant.

We note that our conclusions are different from other recent works, such as [He et al. \(2018\)](#). This is due to the focus on different scales ([He et al.](#) concentrated on small and highly nonlinear scales), and reflects the strong scale-dependence of the behaviours in some MG models.

### 7.3 Future perspective

Our analysis on the eBOSS quasar samples identified several important systematics that will be important for upcoming surveys. The catastrophic redshift failure effect that has an impact on the cosmic growth rate needs to be included in the modelling of the two-point statistics. We show that during full shape analysis the compression of the cosmological information in terms of the AP parameter is an approximation. For the current statistical precision we can achieve using the quasar sample, the accuracy is allowed, but it requires a rigorous treatment in future. Our improved model on the fibre collision effect shows improved accuracy in extracting the cosmological parameter. However, it is inefficient in dealing with the samples with higher projected number density. More exact treatment of the fibre collision effects needs to be integrated in the future work.

Our results on the mocks built on simulation including modified gravity suggest that, with the upcoming galaxy surveys such as DESI, 4MOST and EUCLID, there is a realistic possibility to put constraints on the growth rate of matter and make distinctions between certain MG models and GR. Such studies will potentially benefit from combining with



cosmological data that probe different regimes (e.g., environments), scales and special theoretical properties of the models. For instance, the expected error bar of these future galaxy surveys would help to put tighter constraints on the linear growth rate and help to make a clearer distinction between different gravity models. Also, to be more realistic, it will be useful to test the constraining power of RSD using different tracers and number densities, and include systematic effects such as survey geometric and masks, galaxy redshift distribution and evolution, incompleteness due to fibre collisions and observing conditions, and so on. It is also interesting to study if including higher-order statistics, such as the 3-point correlation function or bispectrum in redshift space, can improve the constraining power of the surveys. In this analysis, we did not include the neutrino effects, which can be degenerate with the effect introduced by modified gravity. Our modelling of two-point statistics does not include the modified gravity effects, which could be included the perturbative calculation. These possibilities will be left for future work.

The analyses presented in this work are still limited to the two-point statistics. It will also be interesting to use the morphology of cosmic large-scale structure such as the Minkowski functionals. The model of the two-point statistics is restricted only upto mildly non-linear scales. Although the perturbative calculation of the gravitational nonlinear evolution has been shown to accurately predict the nonlinear power spectrum at smaller scales, it is theoretically hard to find a good model for the FoG effect below the mildly nonlinear regimes. The information encoded in the small scales is crucial in extracting the constraints on the total neutrino mass, as one of the main goals for the future surveys.



## References

- Abbott B. P., et al., 2016a, *Phys. Rev. Lett.*, 116, 061102
- Abbott B., et al., 2016b, *Physical Review Letters*, 116
- Alam S., et al., 2017, *MNRAS*, 470, 2617–2652
- Alam S., et al., 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for the eBOSS Emission Line Galaxy Sample
- Albareti F. D., et al., 2017, *ApJS*, 233, 25
- Alcock C., Paczynski B., 1979, *Nature*, 281, 358
- Amendola L., et al., 2018, *Living Reviews in Relativity*, 21, 2
- Anderson T., 2003, *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics, Wiley, <https://books.google.de/books?id=Cmm9QgAACAAJ>
- Anderson L., et al., 2012, *MNRAS*, 427, 3435
- Anderson L., et al., 2014, *MNRAS*, 439, 83
- Ata M., et al., 2017, *MNRAS*, p. stx2630
- Barreira A., Sánchez A. G., Schmidt F., 2016, *Phys. Rev.*, D94, 084022
- Barreira A., Krause E., Schmidt F., 2018, *J. Cosmology Astropart. Phys.*, 6, 015
- Bautista J. E., et al., 2017, *Astronomy & Astrophysics*, 603, A12
- Bautista J., et al., 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey Luminous Red Galaxy Sample: BAO and growth rate of structure from the anisotropic correlation function between redshift 0.6 and 1, In preparation
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, *Astrophys. J.*, 762, 109
- Bel J., Pezzotta A., Carbone C., Sefusatti E., Guzzo L., 2019, *A&A*, 622, A109
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
- Bernardeau F., Crocce M., Scoccimarro R., 2012, *Phys. Rev. D*, 85, 123519
- Beutler F., et al., 2011, *MNRAS*, 416, 3017

- Bianchi D., Percival W. J., 2017, *MNRAS*, 472, 1106
- Blanton M. R., et al., 2017, *AJ*, 154, 28
- Blomqvist M., et al., 2019, arXiv e-prints, p. arXiv:1904.03430
- Blot L., et al., 2018, preprint, (arXiv:1806.09497)
- Bose B., Koyama K., 2016, *J. Cosmology Astropart. Phys.*, 2016, 032
- Bose S., Hellwing W. A., Li B., 2015, *J. Cosmology Astropart. Phys.*, 2015, 034
- Bose B., Koyama K., Lewandowski M., Vernizzi F., Winther H. A., 2018, *J. Cosmology Astropart. Phys.*, 2018, 063
- Bovy J., et al., 2012, *ApJ*, 749, 41
- Brandenberger R., Feldman H., Mukhanov V., 1993, *Classical and Quantum Theory of Perturbations in Inflationary Universe Models* (arXiv:astro-ph/9307016)
- Brax P., van de Bruck C., Davis A.-C., Shaw D., 2010, *Phys. Rev.*, D82, 063519
- Brax P., Davis A.-C., Li B., Winther H. A., 2012a, *Phys. Rev.*, D86, 044015
- Brax P., Davis A.-C., Li B., 2012b, *Phys. Lett.*, B715, 38
- Burden A., Percival W. J., Manera M., Cuesta A. J., Vargas Magana M., Ho S., 2014, *MNRAS*, 445, 3152–3168
- Burden A., Percival W. J., Howlett C., 2015, *MNRAS*, 453, 456–468
- Carbone C., Petkova M., Dolag K., 2016, *J. Cosmology Astropart. Phys.*, 2016, 034
- Carlson J., Reid B., White M., 2013, *MNRAS*, 429, 1674
- Carter P., Beutler F., Percival W. J., DeRose J., Wechsler R. H., Zhao C., 2019, *The Impact of the Fiducial Cosmology Assumption on BAO Cosmological Parameter Inference* (arXiv:1906.03035)
- Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, 85, 083509
- Chuang C.-H., Kitauro F.-S., Prada F., Zhao C., Yepes G., 2015, *MNRAS*, 446, 2621
- Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, *Phys. Rep.*, 513, 1
- Clowe D., Gonzalez A., Markevitch M., 2004, *ApJ*, 604, 596–603
- Cole S., et al., 2005, *MNRAS*, 362, 505–534
- Collaboration L. S., et al., 2009, *LSST Science Book, Version 2.0* (arXiv:0912.0201)
- Collaboration P., et al., 2016, *A&A*, 594, A13
- Collaboration P., et al., 2018, *Planck 2018 results. VI. Cosmological parameters* (arXiv:1807.06209)

- Colless M., et al., 2001, MNRAS, 328, 1039
- Cooray A., Sheth R., 2002, Physics Reports, 372, 1–129
- Creminelli P., Noreña J., Simonović M., Vernizzi F., 2013, Journal of Cosmology and Astroparticle Physics, 2013, 025–025
- Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519
- Crocce M., Scoccimarro R., 2008, Phys. Rev. D, 77, 023533
- Crocce M., Scoccimarro R., Blas D., "in prep", Galileon invariant renormalized perturbation theory
- Crocce M., Scoccimarro R., Bernardeau F., 2012, MNRAS, 427, 2537
- DESI Collaboration et al., 2016a, arXiv e-prints, p. arXiv:1611.00036
- DESI Collaboration et al., 2016b, arXiv e-prints, p. arXiv:1611.00037
- Dawson K. S., et al., 2013, AJ, 145, 10
- Dawson K. S., et al., 2016, AJ, 151, 44
- Deffayet C., Esposito-Farèse G., Vikman A., 2009, Physical Review D, 79
- Deffayet C., Gao X., Steer D. A., Zahariade G., 2011, Physical Review D, 84
- Desjacques V., Seljak U., 2010, Classical and Quantum Gravity, 27, 124011
- Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
- Desmond H., Ferreira P. G., Lavaux G., Jasche J., 2018, MNRAS, 483, L64–L68
- Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
- Dvali G., Gabadadze G., Porrati M., 2000, Physics Letters B, 485, 208–214
- Efstathiou G., et al., 2002, MNRAS, 330, L29
- Eftekharzadeh S., Myers A. D., Kourkchi E., 2019, MNRAS, 486, 274
- Eggemeier A., Scoccimarro R., Smith R. E., 2019, Phys. Rev. D, 99, 123514
- Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
- Eisenstein D. J., et al., 2005, ApJ, 633, 560
- Fang W., Wang S., Hu W., Haiman Z., Hui L., May M., 2008, Phys. Rev. D, 78, 103509
- Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
- Finkbeiner D. P., et al., 2016, ApJ, 822, 66
- Friedmann A., 1924, Zeitschrift für Physik, 21, 326

- Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
- Gil-Marín H., et al., 2018, MNRAS, 477, 1604
- Gil-Marín H., et al., 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0, In preparation
- Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, MNRAS, 457, 1577
- Grieb J. N., et al., 2017, MNRAS, 467, 2085
- Gunn J. E., et al., 2006, AJ, 131, 2332
- Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodríguez-Torres S., 2017, MNRAS, p. stx185
- Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, The Astronomical Journal, 156, 160
- Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
- Hawkins E., et al., 2003, MNRAS, 346, 78–96
- He J.-h., Guzzo L., Li B., Baugh C. M., 2018, Nature Astronomy,
- Heitmann K., et al., 2019, arXiv e-prints, p. arXiv:1904.11970
- Hernández-Aguayo C., Hou J., Li B., Baugh C. M., Sánchez A. G., 2019, MNRAS, 485, 2194
- Hewett P. C., Wild V., 2010, MNRAS, 405, 2302
- Hinshaw G., et al., 2013, The Astrophysical Journal Supplement Series, 208, 19
- Hinterbichler K., Khoury J., 2010, Phys. Rev. Lett., 104, 231301
- Hoffman K., Kunze R., 1971, Linear Algebra. Prentice-Hall Mathematics Series, N.J., Prentice-Hall, <https://books.google.de/books?id=I4kQAQAIAAJ>
- Horndeski G. W., 1974, International Journal of Theoretical Physics, 10, 363
- Hou J., et al., 2018, MNRAS, 480, 2521
- Hou J., et al., 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey Quasar Sample: BAO and RSD measurements from anisotropic clustering analysis in configuration space between redshift 0.8 and 2.2, In preparation
- Howlett C., Ross A. J., Samushia L., Percival W. J., Manera M., 2015, MNRAS, 449, 848–866
- Hu W., Sawicki I., 2007, Phys. Rev., D76, 064004

- Hubble E., 1929, *Proceedings of the National Academy of Sciences*, 15, 168
- Huchra J., Davis M., Latham D., Tonry J., 1983, *ApJS*, 52, 89
- Hui L., Parfrey K. P., 2008, *Phys. Rev. D*, 77, 043527
- Jackson J. C., 1972, *MNRAS*, 156, 1P
- Joyce A., Jain B., Khoury J., Trodden M., 2015, *Physics Reports*, 568, 1
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kazin E. A., Sánchez A. G., Blanton M. R., 2012, *MNRAS*, 419, 3223
- Kazin E. A., et al., 2014, *MNRAS*, 441, 3524
- Khoury J., Weltman A., 2004, *Phys. Rev. D*, 69, 044026
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kobayashi T., Yamaguchi M., Yokoyama J., 2011, *Progress of Theoretical Physics*, 126, 511–529
- Kobayashi Y., Nishimichi T., Takada M., Takahashi R., 2019, *arXiv e-prints*, p. arXiv:1907.08515
- Komatsu E., et al., 2011, *ApJS*, 192, 18
- Koyama K., 2016, *Reports on Progress in Physics*, 79, 046902
- Koyama K., Silva F. P., 2007, *Phys. Rev. D*, 75, 084040
- Lahav O., et al., 2002, *Mon. Not. Roy. Astron. Soc.*, 333, 961
- Laureijs R., et al., 2011, *arXiv e-prints*, p. arXiv:1110.3193
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Li B., Zhao G.-B., Teyssier R., Koyama K., 2012, *J. Cosmology Astropart. Phys.*, 2012, 051
- Li B., Zhao G.-B., Koyama K., 2013, *J. Cosmology Astropart. Phys.*, 2013, 023
- Linder E. V., Cahn R. N., 2007, *Astroparticle Physics*, 28, 481
- Lippich M., et al., 2019, *MNRAS*, 482, 1786
- Lyke B., Higley A. N., McLane J. N., et al. 2020, *The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release*, In preparation
- MINOS Collaboration et al., 2014, *arXiv e-prints*, p. arXiv:1403.0867
- Maartens R., Majerotto E., 2006, *Phys. Rev. D*, 74, 023004
- Manera M., et al., 2012a, *Mon. Not. Roy. Astron. Soc.*, 428, 1036

- Manera M., et al., 2012b, *Mon. Not. Roy. Astron. Soc.*, 428, 1036
- Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, *ApJ*, 606, 819–824
- Matsubara T., 2011, *Phys. Rev. D*, 83, 083518
- Mead A. J., Peacock J. A., 2014a, *MNRAS*, 440, 1233
- Mead A. J., Peacock J. A., 2014b, *MNRAS*, 445, 3453
- Mohammad F. G., et al., 2018, *Astronomy & Astrophysics*, 619, A17
- Mohammad F., Percival W. J., et al. 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Pairwise-Inverse-Probability Correction for Fiber Collisions in Clustering Measurements, In preparation
- Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, *MNRAS*, 433, 2389–2402
- Mukhanov V. F., Feldman H. A., Brandenberger R. H., 1992, *Phys. Rep.*, 215, 203
- Müller E., eBOSS collaboration et al. 2020, Cosmological Implications of the Final Measurements from the Extended Baryon Oscillation Spectroscopic Survey, In preparation
- Myers A. D., et al., 2015, *ApJS*, 221, 27
- NOvA Collaboration et al., 2017, arXiv e-prints, p. arXiv:1701.05891
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *Astrophys. J.*, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *Astrophys. J.*, 490, 493
- Neveux R., Burtin E., et al. 2020, The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey Quasar Sample: BAO and RSD measurements from anisotropic power spectrum between redshift 0.8 and 2.2, In preparation
- Nishimichi T., Bernardeau F., Taruya A., 2017, *Phys. Rev. D*, 96, 123515
- Oort J. H., 1932, *Bull. Astron. Inst. Netherlands*, 6, 289
- Padmanabhan N., et al., 2008, *ApJ*, 674, 1217
- Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, *MNRAS*, 427, 2132
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144–1156
- Peacock J. A., et al., 2001, *Nature*, 410, 169
- Peloso M., Pietroni M., 2013, *Journal of Cosmology and Astroparticle Physics*, 2013, 031–031
- Percival W. J., et al., 2014, *MNRAS*, 439, 2531



- Perlmutter S., et al., 1999, *ApJ*, 517, 565–586
- Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, *ApJS*, 178, 179
- Putter R. D., Wagner C., Mena O., Licia V., Percival W. J., 2012, *J. Cosmology Astropart. Phys.*, 2012, 019
- Raichoor A., et al., 2020, The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey Emission Line Galaxy Sample: Large-scale Structure Catalogues and Measurement of the isotropic BAO between redshift 0.6 and 1.1
- Reid B. A., White M., 2011, *MNRAS*, 417, 1913
- Reid B. A., et al., 2010, *MNRAS*, 404, 60
- Riess A. G., et al., 1998, *The Astronomical Journal*, 116, 1009–1038
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *ApJ*, 876, 85
- Riotto A., 2002, *Inflation and the Theory of Cosmological Perturbations* (arXiv:hep-ph/0210162)
- Robertson H. P., 1935, *ApJ*, 82, 284
- Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015a, *MNRAS*, 449, 835
- Ross A. J., Percival W. J., Manera M., 2015b, *MNRAS*, 451, 1331
- Ross A. J., et al., 2017, *MNRAS*, 464, 1168
- Ross A., Bautista J., et al. 2020, Large-scale Structure Catalogs for Cosmological Analysis of the completed extended Baryon Oscillation Spectroscopic Survey, In preparation
- Rossi G., Hou J., et al. 2020, N-body Mock Challenge for eBOSS DR16 Galaxy Clustering Measurements, In preparation
- Ruggeri R., et al., 2019, *MNRAS*, 483, 3878
- Sánchez A., 2018, Lecture notes for LSS, In preparation
- Sánchez A., 2020, Let us bury the prehistoric  $h$ : arguments against using  $h^{-1}$ Mpc units in observational cosmology (arXiv:2002.07829)
- Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, *MNRAS*, 366, 189
- Sánchez A. G., et al., 2017a, *MNRAS*, 464, 1640
- Sánchez A. G., et al., 2017b, *MNRAS*, 464, 1493–1501
- Saunders W., et al., 2000, *MNRAS*, 317, 55

- Schneider D. P., et al., 2010, *AJ*, 139, 2360
- Soccimarro R., 2004, *Phys. Rev. D*, 70, 083007
- Seldner M., Siebers B., Groth E. J., Peebles P. J. E., 1977, *AJ*, 82, 249
- Shane C. D., Wirtanen C. A., 1967, *Pub. Lick Obs.*, 22
- Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385
- Slipher V. M., 1914, *Popular Astronomy*, 22, 19
- Smee S. A., et al., 2013, *AJ*, 146, 32
- Smith A., Hou J., et al. 2020, The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for the Quasar Sample, In preparation
- Spergel D., et al., 2013, arXiv e-prints, p. arXiv:1305.5425
- Springel V., 2005, *MNRAS*, 364, 1105
- Tadros H., et al., 1999, *MNRAS*, 305, 527
- Takahashi R., 2008, *Progress of Theoretical Physics*, 120, 549
- Tamone A., et al., 2020, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey ELG sample: Growth rate of structure measurement in configuration space between redshift 0.6 and 1.1., In preparation
- Taruya A., 2016, *Phys. Rev. D*, 94, 023504
- Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
- Taruya A., Bernardeau F., Nishimichi T., Codis S., 2012, *Phys. Rev. D*, 86, 103528
- Taylor A. N., Dye S., Broadhurst T. J., Benitez N., van Kampen E., 1998, *ApJ*, 501, 539
- Teyssier R., 2002, *A&A*, 385, 337
- Tinker J. L., et al., 2012, *ApJ*, 745, 16
- Vainshtein A. I., 1972, *Physics Letters B*, 39, 393
- Verde L., Jimenez R., Kamionkowski M., Matarrese S., 2001, *Monthly Notices of the Royal Astronomical Society*, 325, 412–418
- Walker A. G., 1937, *Proceedings of the London Mathematical Society*, s2-42, 90
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *The Astrophysical Journal*, 652, 71
- White M., Reid B., Chuang C.-H., Tinker J. L., McBride C. K., Prada F., Samushia L., 2015, *MNRAS*, 447, 234

- Wright E. L., et al., 2010, *AJ*, 140, 1868
- Wu P., Yu H., Fu X., 2009, *J. Cosmology Astropart. Phys.*, 2009, 019
- Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2013, *MNRAS*, 431, 2834
- York D. G., et al., 2000, *AJ*, 120, 1579
- Zarrouk P., et al., 2018, *MNRAS*, 477, 1639
- Zhao G.-B., et al., 2019, *MNRAS*, 482, 3497
- Zhao C., Chuang C., et al. 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: one thousand multi-tracer mock catalogues with redshift evolution and systematics for galaxies and quasars of the final data release, In preparation
- Zheng Z., et al., 2005, *Astrophys. J.*, 633, 791
- Zheng Z., Coil A. L., Zehavi I., 2007, *Astrophys. J.*, 667, 760
- Zwicky F., 1933, *Helvetica Physica Acta*, 6, 110
- de Mattia A., Ruhlmann-Kleider V., 2019, *J. Cosmology Astropart. Phys.*, 2019, 036–036
- de Mattia A., et al., 2020, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey emission line galaxy sample: measurement of the BAO and growth rate of structure from the anisotropic power spectrum between redshift 0.6 and 1.1, In preparation
- de Sainte Agathe V., et al., 2019, arXiv e-prints, p. arXiv:1904.03400
- du Mas des Bourboux H., et al., 2017, *Astronomy & Astrophysics*, 608, A130
- du Mas des Bourboux H., et al., 2020, The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyman-alpha Forests, In preparation



## Appendix A

### BAO reconstruction of the QSO catalog

As we showed in Section 3.1 the nonlinear gravitational evolution leads to shifted and damped BAO signal. Since the signal degradation is sourced by the gravitational evolution, one can in principle partially restore the signal by estimating the gravitational potential from the tracer density. Following Padmanabhan et al. (2012), the displacement field,  $\Psi(\mathbf{q}, t)$ , that links the positions in the initial real space to the one in the evolved redshift space can be obtained by solving the following equation

$$\nabla \cdot \Psi + \frac{f}{b} \nabla \cdot (\Psi \cdot \hat{\mathbf{r}}) \hat{\mathbf{r}} = -\frac{\delta_g}{b}, \quad (\text{A.1})$$

where  $\hat{\mathbf{r}}$  is the line of sight direction,  $b$  is the linear large-scale structure bias and  $f$  is the linear growth rate. We used a method proposed by Burden et al. (2014, 2015) to perform reconstruction in Fourier space<sup>1</sup>, so that the displacement solved in the Fourier space reads

$$\Psi = \text{IFFT} \left[ -\frac{i\mathbf{k}\delta(\mathbf{k})}{k^2 b} \right] - \frac{f}{b+f} \left\{ \text{IFFT} \left[ -\frac{i\mathbf{k}\delta(\mathbf{k})}{k^2 b} \right] \cdot \hat{\mathbf{r}} \right\} \hat{\mathbf{r}}, \quad (\text{A.2})$$

where IFFT denotes inverse Fourier transform.<sup>2</sup>

We test the reconstruction on the OUTERRIM simulation. Fig. A.1 shows a projection along the x-axis of the simulation box with a slice thickness of  $50 h^{-1} \text{Mpc}$ . The background in this plot shows the density field before (left) and after 1-iteration of reconstruction (right). We can clearly see that the density contrast is reduced after the BAO reconstruction, as the reconstruction removes the non-linearity. The arrows in Fig. A.1 shows the displacement field. The smoothing scale in this plot is set to be  $ds = 20 h^{-1} \text{Mpc}$ .

<sup>1</sup>the code is based on [https://github.com/julianbautista/eBoss\\_clustering](https://github.com/julianbautista/eBoss_clustering)

<sup>2</sup>Note in Burden et al. (2014) the coefficient in the second term of the denominator is mistaken for  $\frac{f}{1+f}$ .

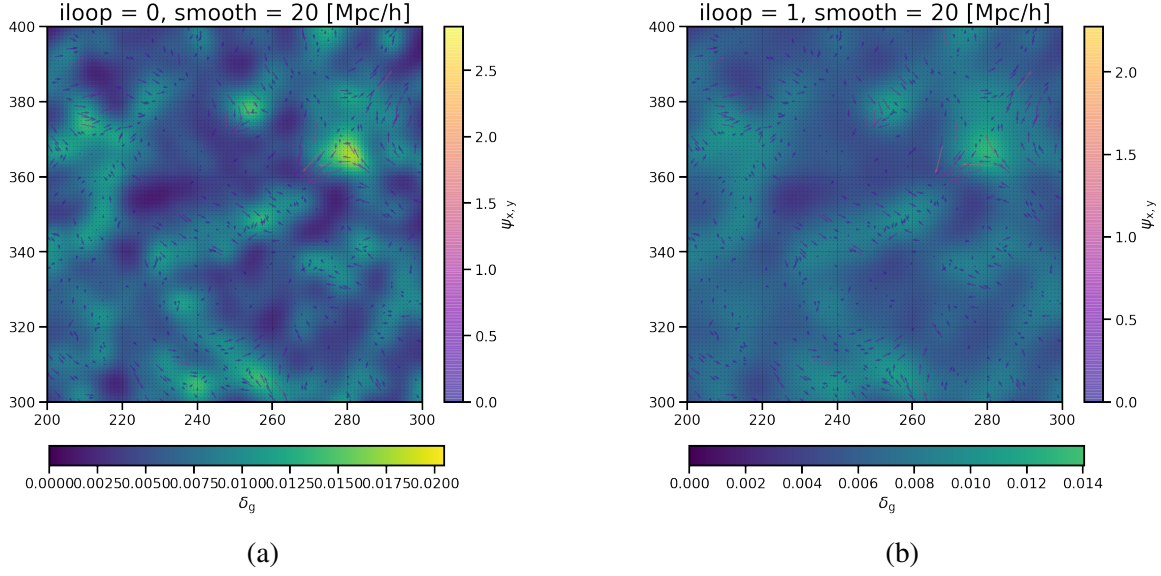


Fig. A.1 A slice of the BAO reconstructed density field projected along the x-axis of the simulation box. Background: density contrast pre-reconstruction (left) and post-reconstruction (right). Arrows: displacement field.

Fig A.2 shows the measurement of the two point correlation function using the OuterRim mock with different smoothing scales, ranging from  $20 h^{-1}\text{Mpc}$  to  $50 h^{-1}\text{Mpc}$  with 3 iteration steps. The orange solid line is the mean of the mocks pre-reconstruction. In this plot, firstly we can see that reconstruction undo the RSD boost that reduces the monopole amplitude at small scales and enhance the BAO contrast. Secondly, reconstruction reduces the amplitude of the quadrupole but at the same time it becomes more positive as the smoothing length decreases.

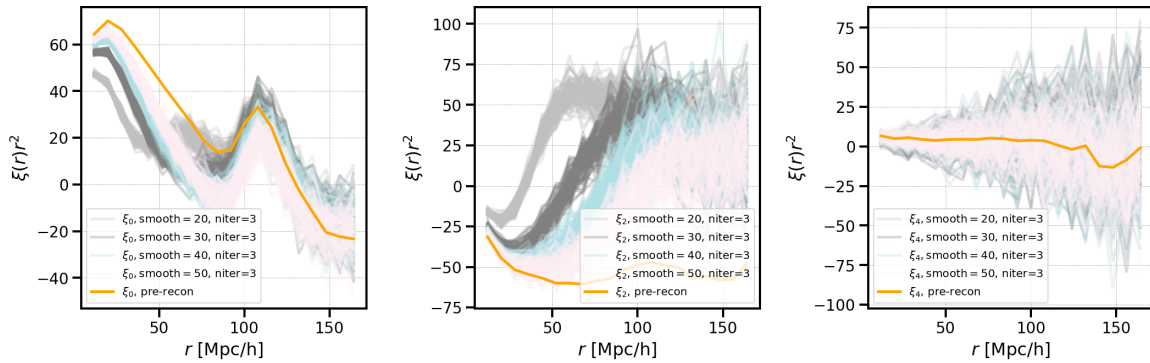


Fig. A.2 Two point correlation for reconstruction with different smoothing scales. The orange curve is the mean of the mocks pre-reconstruction.

Fig. A.3 shows the measurement of the three Legendre multipoles after BAO reconstruction applied to the DR16 eBOSS quasar sample. On the real data, we observe very similar behaviour on the three multipoles as compared to the OUTERRIM mocks that, the quadrupole becomes positive after the reconstruction is applied. The reason is likely due to the high shot noise in the QSO sample. In the first step of the reconstruction, a Gaussian-smoothed density field is estimated from the discrete tracers, at the same time, the uncorrelated shot noise is also smoothed and leads to a contaminated estimation of the density field. This is not a big problem when the number density of the tracers is high. However, the number density for quasars is as low as  $n(z) \sim 10^{-5} [h^{-1}\text{Mpc}]^{-3}$ , the conventional BAO reconstruction is highly contaminated by the shot noise. Therefore, despite the enhanced BAO signal in the monopole, we decided not to apply BAO reconstruction to the quasar sample.

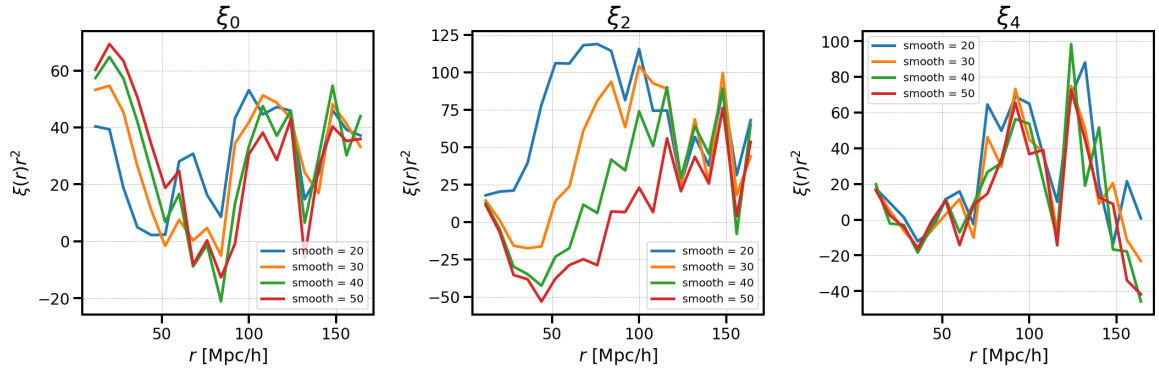


Fig. A.3 Two point correlation for reconstruction with different smoothing scales.





## Appendix B

# Supplementary to the eBOSS clustering analysis

### B.1 Results expressed in alternative basis

In this appendix, we consider the results given in different alternative basis. First, we list the results expressed in the basis of comoving angular diameter  $D_M/r_{\text{drag}}$ <sup>1</sup>, Hubble parameter,  $Hr_{\text{drag}}$  and  $f\sigma_8$ . Second, we list the results expressed in the basis of comoving angular diameter  $D_M/r_{\text{drag}}$ , Hubble distance,  $D_H/r_{\text{drag}}$  and  $f\sigma_{12}$ .

#### B.1.1 $D_M/r_{\text{drag}}$ - $Hr_{\text{drag}}$ - $f\sigma_8$ basis

##### BAO-only fits

We convert the BAO  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  results to constraints on the comoving angular diameter and Hubble parameter with:

$$D_M(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.82(80)(21), \quad (\text{B.1})$$

$$H(z_{\text{eff}} = 1.48)r_{\text{drag}} = 2.267(93)(25) \times 10^4 \text{ km/s}. \quad (\text{B.2})$$

The first parentheses denotes the statistical error, which is rescaled by  $\sqrt{M} = 1.010$ , the second parentheses denotes the error inferred from the mock challenge.

---

<sup>1</sup>The comoving angular diameter is defined as  $D_M = \int_0^z \frac{cdz'}{H(z')}$ .

### Full-shape analysis in configuration space

$$D_M(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.66(84)(22)(10), \quad (\text{B.3})$$

$$H(z_{\text{eff}} = 1.48)r_{\text{drag}} = 2.289(85)(26)(11) \times 10^4 \text{ km/s}, \quad (\text{B.4})$$

$$f\sigma_8(z_{\text{eff}} = 1.48) = 0.439(46)(13)(4), \quad (\text{B.5})$$

where the first parentheses denotes the statistical error, which is rescaled by  $\sqrt{M} = 1.036$ , the second parentheses denotes the modelling systematics inferred from the OUTERIM mock challenge, the third parentheses denotes the observational systematics inferred from the EZMOCKS. The covariance matrix is given by

$$\mathbf{C}_\lambda = \begin{pmatrix} D_M/r_{\text{drag}} & H r_{\text{drag}} & f\sigma_8 \\ 7.709 \times 10^{-1} & 9.780 \times 10^1 & 1.750 \times 10^{-2} \\ - & 8.007 \times 10^5 & 1.085 \times 10^1 \\ - & - & 2.308 \times 10^{-3} \end{pmatrix} \quad (\text{B.6})$$

### Combined full-shape analysis

$$D_M^c(z_{\text{eff}} = 1.48)/r_{\text{drag}} = 30.21 \pm 0.79, \quad (\text{B.7})$$

$$H^c(z_{\text{eff}} = 1.48)r_{\text{drag}} = 2.255 \pm 0.790 \times 10^4 \text{ km/s}, \quad (\text{B.8})$$

$$f\sigma_8^c(z_{\text{eff}} = 1.48) = 0.459 \pm 0.045, \quad (\text{B.9})$$

where the errors include the statistical and systematic uncertainties (inferred from OUTERIM mock challenge and EZMOCKS). The covariance matrix for the combined data is given by

$$\mathbf{C}_\lambda^c = \begin{pmatrix} D_M/r_{\text{drag}} & H r_{\text{drag}} & f\sigma_8 \\ 6.222 \times 10^{-1} & -4.374 \times 10^1 & 2.223 \times 10^{-2} \\ - & 6.216 \times 10^5 & 1.191 \times 10^1 \\ - & - & 2.010 \times 10^{-3} \end{pmatrix} \quad (\text{B.10})$$

## B.2 Parameter settings for the OuterRim mock challenge

Table B.1 The HOD settings for the non-blind mocks. The HODs are combinations of Gaussian, smooth-, sharp- step function, top hat and power law (PL). In the case of ”+” sign, the central galaxies and the satellite galaxies are placed using two different recipes.

	<b>mock0</b>	<b>mock1</b>	<b>mock2</b>	<b>mock3</b>	<b>mock4</b>	<b>mock5</b>
HOD	Smooth Step + PL	Gaussian	Top Hat + PL	Gaussian +PL	Smooth Step + PL	Sharp Step + PL
Satellite Position	Particle	NFW	NFW	Particle	NFW	NFW
Satellite Fraction %	19	7	60	21	8	17
	<b>mock6</b>	<b>mock7</b>	<b>mock8</b>	<b>mock9</b>	<b>mock10</b>	<b>mock11</b>
HOD	Gaussian + PL	Top Hat + PL	Gaussian	Sharp Step + PL	Sharp Step + PL	Gaussian
Satellite Position	NFW	Particle	NFW	NFW	Particle	NFW
Satellite Fraction %	56	24	100	42	0.2	10
	<b>mock12</b>	<b>mock13</b>	<b>mock14</b>	<b>mock15</b>	<b>mock16</b>	<b>mock17</b>
HOD	Gaussian + PL	Smooth Step + PL	Top Hat + PL	Gaussian	Top Hat + PL	Sharp Step + PL
Satellite Position	NFW	NFW	Particle	Particle	NFW	NFW
Satellite Fraction %	5	73	17	50	12	4
	<b>mock18</b>	<b>mock19</b>				
HOD	Smooth Step + PL	Gaussian + PL				
Satellite Position	NFW	Particle				
Satellite Fraction %	36	7				

Table B.2 Parameters for the blind OuterRim mocks.  $z_{\text{orig}}$  is the redshift of the original snapshot which is rescaled to the new cosmology at  $z = 1.433$ .  $s_m$  and  $s_v$  are the factors used to scale positions, masses and velocities, respectively.

	$\Omega_b$	$\Omega_{\text{cdm}}$	$h$	$\sigma_8$	$n_s$	$z_{\text{orig}}$	$s$	$s_m$	$s_v$
Mock	0.0448	0.2200	0.7100	0.8000	0.9630	1.433	1.0000	1.0000	1.0000
cosmo0	0.0461	0.2205	0.7228	0.7742	0.9628	1.494	0.9827	0.9555	0.9734
cosmo1	0.0426	0.2360	0.6967	0.7981	0.9384	1.494	1.0175	1.1085	1.0308
cosmo2	0.0410	0.2331	0.7405	0.7380	0.9594	1.494	0.9087	0.7766	0.9129
cosmo3	0.0447	0.2408	0.6882	0.7903	0.9436	1.494	0.9950	1.0624	1.0208
cosmo4	0.0467	0.2202	0.6964	0.7991	0.9815	1.376	0.9628	0.8997	0.9788
cosmo5	0.0382	0.1973	0.7197	0.8526	0.9606	1.376	1.1049	1.1992	1.0535
cosmo6	0.0541	0.2295	0.7275	0.7910	0.9671	1.376	0.9327	0.8688	0.9779
cosmo7	0.0475	0.1844	0.7239	0.7783	1.0280	1.376	0.9603	0.7756	0.9086

## Acknowledgements

I would like to acknowledge my supervisor Ariel Sánchez for guiding me into the topic of cosmology. Thank you that I had fun and freedom of exploring physics during my PhD time. I acknowledge Prof. Ralf Bender for the support of my PhD work. I thank my colleges, Daniel Farrow, Martha Lippich, Salvador Salazar-Albornoz, and Agne Semenaite for sharing your time with me and having productive discussions. I thank my lunch time group members, Alessandro Ballone, Kianush Merghan, and Martin Zintl for inspiring philosophical and political thoughts during the breaks. I thank the nice people I met at MPE, Manuel Behrendt, Matias Blana, Marcello Crosta, Fabrizio Finozzi, Mattias Kluge, Guangxin Li, Roberto Saglia, and Wolfgang Schulter. I thank the people I met at Durham university, César Hernández-aguayo, Prof. Baojiu Li, and Prof. Carlton Baugh for hosting me and I learnt really a lot during the visit. I also thank my colleges at eBOSS collaboration, we discuss, we doubt, we fight, we learn, we reconcile. The song of “eBOSS” is a wonderful memory under the night sky.

There’re many more people I want to thank, I do not write down your special name, but you know that your name is listed in this space “ ”, and you know that I do remember. I enjoy my decision switching the topic to cosmology, people who I met during my PhD time help me to sort the universal net, the one I wish to complete on the course of experiencing the sequence of events.

I would like to thank the great support from Hao Ding. You know, the relation is hidden in the unknown out-burst of the flow of life and the dust of the red giant. Sorry to disappoint you, my dear grand parents, that I will not grow into a hero that save the world with supremacy. Rather, I am a cosmic voyager. Finally, I would like to thank my parents, thank you for your trust that I’m not randomly left to a lonely planet.



